Developing a mass spectrometric method for P-III-NP to detect rhGH administration

Moncrieffe, Danielle Analise

Awarding institution:
King's College London

Developing a mass spectrometric method for P-III-NP to detect rhGH administration

Ph.D. thesis

In partial fulfilment of the requirement for the award of Ph.D. in Analytical Science

2018

Supervisors:

Prof. David Cowan

Dr. Mark Parkin

Declaration:

I confirm that the work presented in this thesis is my own and all references are cited accordingly.

Abstract

Procollagen III amino-terminal propeptide (P-III-NP), a biomarker for detecting growth hormone administration, is currently measured by immunoassays in the absence of international reference material. The main aim of this research is to demonstrate whether liquid chromatography tandem mass spectrometry (LC-MS) can be used to quantify P-III-NP in blood.

Low serum concentration (1-5 ng/mL, 25-125 pM) and medium protein size of P-III-NP (~42 kDa) complicates method development, in that there is a requirement for very sensitive LC-MS methods. To meet this, a digest approach with trypsin has been adopted. By LC-HRAMS (high resolution accurate MS in tandem mode) analysis of digested bovine P-III-NP material, post translational modifications (PTMs) of some amino acid residues have been identified through the partial *de novo* sequencing of peptides. Once assessed for stability, peptides T1 and T5 were synthesized and used to develop sensitive nano- and micro-flow LC-MS methods on a highly sensitive triple quadrupole MS. By both approaches, adequate sensitivity of peptides (with no matrix) has been achieved to accommodate analysis of P-III-NP at basal levels.

However, despite the sensitivity of these developed methods, the presence of albumin (HSA, 35-50 mg/mL, 5.3-7.75 x 10^{-1} mM) in the sample matrix prevents analysis by LC-MS. Through investigation, it has been determined that HSA concentrations > 20 µg/mL results in the complete ion suppression of P-III-NP peptides (100 pM) in the MS. Hence, HSA depletion ≥ 99.6 % is necessary to produce samples suitable for LC-MS. Investigations into common proteomic HSA depletion methods have shown that double depletion methods, i.e. protein precipitation (ppt) then molecular weight cut off (MWCO) filters or double MWCO, should be effective in depleting HSA sufficiently from serum. However, further investigations have highlighted that the co-precipitation of P-III-NP with HSA and the poor reproducibility of P-III-NP recovery from MWCO limit the application of these approaches for preparing our serum samples for LC-MS.

With the recent (limited) release of an ELISA immunoassay kit targeting P-III-NP, immunocapture of serum P-III-NP prior to trypsin digestion and LC-MS has enabled the identification of peptides from endogenous sources. Using the T5 peptide, the *h*P-III-NP concentration in a normal human serum sample was semi-quantitatively determined as ~2 ng/mL. For absolute quantification, however, suitable internal standards are needed. Isotopically heavy labelled *h*T1 and T5 peptide variants have been suggested, for which synthesis and verification has commenced and is presented.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

List of Abbreviations

A	*Ala*	Alanine
aa		amino acid
ACN		Acetonitrile
ACTH		Adrenocorticotropic hormone
ALS		Acid labile subunit
AGC		Automatic gain control
amu		arbitrary mass unit
BLAST		Basic local alignment search tool
BMI		Body mass index
C	*Cys*	Cysteine
cDNA		Coding DNA
CID		Collision induced dissociation
D	*Asp*	Aspartic acid
DIC		*N,N'*-Diisopropylcarbodiimide
DIPEA		*N,N*-Diisopropylethylamine trihydrofluoride
DNA		Deoxyribonucleic acid
DTT		Dithiothritol
E	*Glu*	Glutamic acid (Glutamate)
ELISA		Enzyme-linked immunosorbence assay
ESI		Electrospray ionisation
F	*Phe*	Phenylalanine
FWHM		Full width half maximum
G	*Gly*	Glycine
GHD		Growth hormone deficiency
GHRH		Growth hormone releasing hormone
GHRP		Growth hormone releasing peptide
GalNac		*N*-acetylgalactosamine
GlcNac		*N*-acetylglucosamine
H	*His*	Histidine
HATU		1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluoro-phosphate
HESI		Heated electrospray ionisation
HMW		High molecular weight
HRAMS		High resolution accurate mass spectrometry
HSA		Human serum albumin
*hydroxy*P		Hydroxyproline
I	*Ile*	Isoleucine

IDA		Iodoacetamide
ICTP		procollagen I carboxyl-terminal propeptide
IGF-I		Insulin growth factor I
IGF-BP		Insulin growth factor binding proteins
K	*Lys*	Lysine
L	*Leu*	Leucine
LAP		Low abundant protein
LC		Liquid chromatography
LOD		Limit of detection
LMW		Low molecular weight
M	*Met*	Methionine
MALDI		Matrix assisted laser desorption ionisation
MS		Mass spectrometry (spectrometer)
MW		molecular weight
MWCO		Molecular weight cut off
m/z		Mass to charge ratio
N	*Asn*	Asparagine
nCE		normalised collision energy
P	*Pro*	Proline
pI		Isoelectric point
P-III-NP		Procollagen III amino terminal propeptide
ppm		Parts per million
PTM		Post translational modification
*pyro*E		Pyroglutamate
Q	*Gln*	Glutamine
R	*Arg*	Arginine
RIA		Radioimmunoassay
rhGH		Recombinant human growth hormone
RNA		Ribonucleic acid
S	*Ser*	Serine
SPPS		Solid phase peptide synthesis
T	*Thr*	Threonine
ToF		Time of flight
V	*Val*	Valine
W	*Trp*	Trytophan
WADA		World Anti-doping Agency
Y	*Tyr*	Tyrosine

Acknowledgements

I would especially like to thank my supervisors, Prof. David Cowan and Dr. Mark Parkin, for their continuous support and guidance. You both have set an impeccable standard of scientific excellence and nurtured in me a desire to strive for it. I would also like to thank the Partnership for Clean Competition for funding and supporting my project throughout. I am grateful to Dr. Alessandro Musenga, who was involved during the first year of this Ph.D. as a supervisor, giving invaluable support.

Thanks to Christiaan Bartlett and Dr. Vincenzo Abbate (and his research group) for their technical expertise. To Alison Woffendin, thanks for the tremendous amount of time you spent proof reading. Thanks to all the staff at the Drug Control Centre, you have always been supportive and pushed me towards reaching this goal. To the members of King's Forensics and the Department of Analytical, Environmental and Forensic Sciences, your words of encouragement have bolstered me throughout this process, especially during the writing up stage.

To my family and friends, thank you for a lifetime of love and support. This journey would not have been the same without the care and camaraderie of the Tough Tough mudders, drinkers, dancers, swimmers, runners and singers. You guys have made the time doing this Ph.D. so enjoyable, I might just do another one...

Chapter 1 *: General Introduction*

Human growth hormone (hGH), due to its anabolic and lipolytic properties is misused in sport, despite it being prohibited by the World Anti-Doping Agency (WADA) [1]. Intended for use for the therapeutic treatment of GH-related hypopituitarism, access to hGH is tightly controlled. Nevertheless, athletes have been able to obtain hGH for doping, in either its recombinant human (rhGH) or pituitary-derived "cadaveric" form (both have identical amino acid (aa) sequences), through the black market. Abuse among athletes is significant, especially in the bodybuilding and strength athletic community [2]. Besides its ergogenic effects and aid in injury recovery, athletes chose this drug for doping over others that give a similar effect due to the difficulties associated with detection of hGH by analytical testing methods [3]. Aware of this, the anti-doping community has been engaged in research before 2000 to develop an appropriate approach to detect hGH administration [4, 5]. Resulting from their efforts, several testing methods have been developed for hGH, however, research is on-going to further develop these approaches to improve discrimination between synthetic (rhGH, 22 kDa GH) and pseudo-endogenous (cadaveric GH, identical isoform distribution as endogenous hGH) from natural hGH, in the hopeful attempt of eliminating GH abuse in sport.

1.1 **Growth hormone**

GH (also known as somatotropin) is a protein hormone that stimulates growth and cellular reproduction and regeneration in humans and increases muscle mass and strength whilst reducing body fat. Only human GH is bioactive in humans, as this hormone shows species specificity [6]. The majority of the circulating hGH is made up of the 22 kDa isoform (70 %), which has a half-life of 15-20 minutes. Due to RNA (ribonucleic acid) splicing, 20 kDA (5-10 %) and 17 kDA (4%) bioactive isoforms as well as others in smaller amount exist that have longer half lives than the 22 kDA variant [7-9]. Also, through the formation of oligomers, post-translational modification (PTM, e.g.

acetylation, deamidation or phosphorylation) and peptide fragmentation of the aforementioned isomers, less abundant variants exist in circulation [8, 10].

The main source of circulating hGH is through the pulsative secretion of the somatotrope cells in the anterior pituitary. During pregnancy, females secrete additional hGH in the placenta [11]. In humans the basal range of pituitary GH is between 0.01-1 ng/mL, however after a secretary pulse this range increases to 1-100 ng/mL [3]. These wide hGH reference ranges are associated with the combined effects of age, gender, nutritional state, sleep, body composition, stress, various hormones and physical exercise on GH release. Age is a significant modulator of GH, in that concentrations decrease by ~14 % per decade after age 40 [12]. This is not surprising as a number of physiological age related conditions (e.g. reduced muscle mass, strength and energy) are similar to that of GH deficient (GHD) adults.

The amount of GH secreted in each pulse is physiologically regulated by the peptidyl agonist and antagonistic actions of three peptides: GH releasing hormone (GHRH), somatostatin and GH releasing peptide (GHRP, Ghrelin) [13]. In the brain, GHRH from the hypothalamus stimulates GH production and secretion, and somatostatin inhibits release without affecting synthesis [14, 15]. GHRP, which is expressed in the stomach, anterior pituitary and hypothalamus, increases the secretion of GH through an associated receptor [12, 16]. GHRP plays a minor role in GH secretion in that, although capable of direct stimulation, it predominantly works synergistically to GHRH [17]. Additional secretion can be achieved through a series of compounds called GH secretagogues, that act as agonists of either GHRH or GHRP to increase GH release.

GH pulses occur every 2-3 h, at varying amplitudes, with the largest pulses normally occurring at night during the slow wave stage of the sleep cycle [18, 19]. Once released, GH interacts with two receptors at different binding sites to start a signalling cascade that results in the generation of insulin-like growth factor-I (IGF-I) [14]. GH inhibits its own secretion, either directly or indirectly, through IGF-I, using a negative

feedback mechanism at either the hypothalamic or pituitary glands; shown in Figure 1-1.

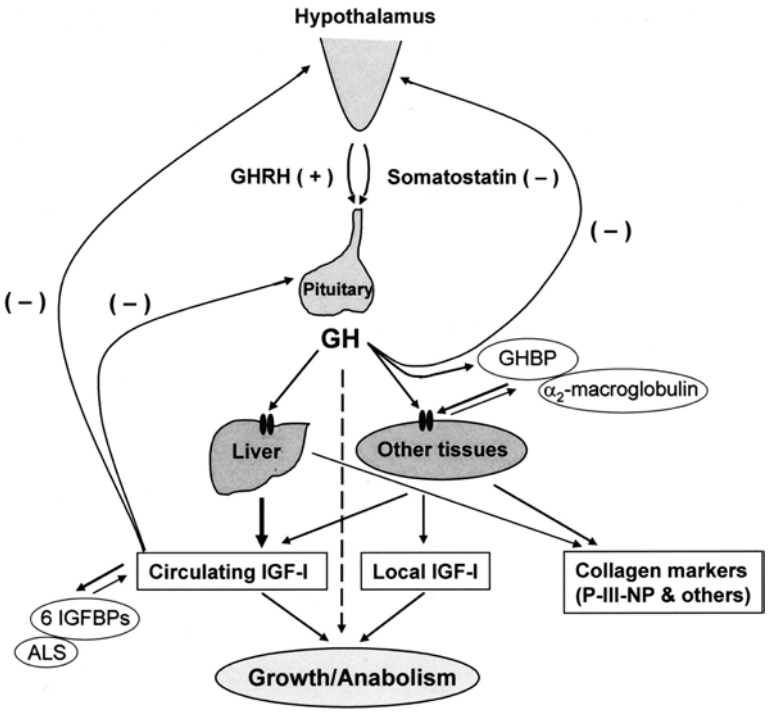


Figure 1.1-1: The GH-IGF-I axis showing the inhibitory (-) and stimulatory (+) actions that regulate hormone secretion. The dashed line indicates that GH has a direct effect on growth and anabolism. In response to GH release, tissues produce IGF-I and the collagen markers, with IGF-I playing a significant role in GH regulation (Image taken from [3] with permission).

On the basis of its lipophilic and anabolic properties, GH is considered to be a powerful metabolic hormone. Through the anabolic actions of GH an increase in protein anabolism and decrease in catabolism in the skeletal muscles occurs, this results in the increase of hypertrophic and hyperplasic growth in muscles, strength and exercise performance [20, 21]. GH also stimulates the formation of bone in excess of its resorption, to lead to a build-up of peak bone mass. These anabolic properties are primarily mediated by circulating IGF-I [21], although GH is also a direct stimulus (shown in Figure 1-1). Both indirectly (through the actions of IGF-I) and directly, GH stimulates lypolysis to increase the concentration of circulating free fatty acids for metabolism in preference to carbohydrates to become the main source of energy in the body [22].

1.1.1 *GH deficiency and treatment*

Continued production of GH by the anterior pituitary glands is needed or GHD results. GHD shows different symptoms with age. In infants and children growth failure is the major symptom. Whilst in adults, although rare, the features of GHD include diminished lean body mass, poor bone density, increased body fat (especially abdominal visceral fat) and increased risk of cerebro- and cardiovascular disease [17]. Hormone replacement treatment in children reverses the symptoms of GHD [23]. In adults, some reversal properties of hypopituitarism have been noted where changes in vigour, ambition and sense of wellbeing have been noted upon drug administration [24]. There has been considerable debate over the performance enhancement properties of GH [25-27]. However, studies have indicated that GH exerts performance benefit to abstinent anabolic steroid users [28, 29].

First extracted and purified (from the pituitary glands) in 1956 [23], cadaveric hGH was used to treat GHD until, in 1985, it was identified as one of the causes of the prion induced Creutzfeldt-Jakob disease and was removed from the market. Being pituitary derived, cadaveric hGH is identical to endogenous circulating hGH, containing the full range of isoforms. Using recombinant DNA technology, a recombinant hGH (rhGH), identical to the 22 kDA GH isoform, was developed for therapeutic use to replace the extinct cadaveric hGH [30, 31].

1.2 **Measurement of hGH**

The clearance of hGH in urine is not proportional to that of plasma hGH; as only 0.1-1 % of the blood concentration is reflected in urine [32]. As such hGH is measured in blood. The direct measurement of hGH is complicated due to the combinational effects of the proteins pulsatile release into circulation, short half-life and the heterogenous nature of its isoforms. Despite this, to detect hGH administration for anti-doping tests direct as well as indirect (biomarker) approaches are used.

1.2.1 ***Direct approach***

In response to the negative feedback regulation of GH, exogenous administration of hGH suppresses the endogenous release of the hormone in the body. If rhGH (recombinant form of 22 kDa GH) is administered, down regulation of the hGH is observed, which offsets the natural 22 kDa/pituitary GH isoform ratio to provide a measurable parameter for the determination of drug administration. This aspect is utilised in anti-doping testing approaches, where immunoassay serum measurements of the 22 kDa, 20 kDa [33], and total hGH are used to detect anomalies in the 22:20 kDa or 22 kDa: total hGH ratios. These measurements are unaffected by sex, age, sport discipline or pathological state [34]. Exercise however can have an effect on measurements, where if done in excess endogenous release of GH is promoted and can normalise the isoform ratios to mask exogenous rhGH administration [35].

The main disadvantage of this approach to measuring the GH isoform ratios is the short detection window associated to the half-life of GH. Even rhGH when injected is cleared rapidly (half lives for subcutaneous, intramuscular and intravenous injections are 4-6 h, 3h, 18 min, respectively), and is often undetectable the morning after injection [36]. Thus, if an athlete stops administration a day before testing, this approach is ineffective in detecting doping. If pituitary-derived GH is administered detection by isoform ratio is not possible, as this exogenous material has identical heterogeneity to endogenous GH. Also, other methods of promoting the effects of GH used by athletes (e.g. IGF-I and GH secretagogoues) are undetectable by measuring GH isoform ratios.

1.2.2 ***Biomarker approach***

Alternatively hGH administration can be detected through the measurement of GH responsive biomarker proteins that have longer half-lives and have more stable serum concentrations. Regardless of the source of hGH or the promoter of its effects, by this approach administration is detectable. Placebo controlled studies in recreational athletes have shown an increase in the serum concentrations of the IGF axis proteins

(IGF-I, IGF binding proteins (IGFBP-1, -2 and -3) and acid labile subunit (ALS)) and specific collagen proteins (osteocalcin, procollagen I and III amino-terminal propeptides (P-I-NP and P-III-NP) and procollagen I carboxyl-terminal propeptide (ICTP)) as a result of hGH administration [37-40]. These studies demonstrate the potential of these proteins as biomarkers for exogenous hGH misuse. Differences in the pharmacodynamic profile of the biomarker proteins result in their considerably longer half-lives (90-500 h) compared to hGH (15-20 min, detailed in Figure 1-2 below) [41]; thus facilitating an extension to the window of detection of hGH administration with this approach.

To associate selectivity to the detection of hGH, the combined monitoring of one marker from each marker group is suggested. As the most responsive markers to hGH, IGF-I and P-III-NP are co-monitored in doping control tests [39, 42]. Both proteins have little diurnal variation and are largely unaffected by exercise relative to hGH intake [43]. Body mass index (BMI) and race contribute little to the variability of these markers [43-45]. Age affects the natural levels of both IGF-I and P-III-NP, in that they are inversely proportional, with maximum levels being observed during pubertal years. For P-III-NP slightly higher concentrations exist in men (average concentration 5.4 ±2.3 ng/mL) compared to women (average concentration 5.1 ±1.5 ng/mL); where the opposite is true for IGF-I [44].

The compliance of an athlete's sample is assessed using a calculated GH-2000 score, based on the natural logarithm of IGF-I and P-III-NP concentrations that are combined in a sex-specific and age discriminant function [46, 47]. The formulae used for calculation (shown below in Equation 1-1 and Equation 1-2) were constructed based on normative data collected from elite athletes (n=314), for which their results are specific to the reagents and assays used to measure the biomarkers [43, 48]. Hence, for each WADA approved assay pair for IGF-I and P-III-NP different sex-specific decision limits exist [47].

Equation 1-1: GH-2000 Biomarker score for males [47]

$$-6.584 + 2.905 \times \ln(P - III - NP) + 2.100 \times \ln(IGF - I) - 101.737/age$$

Equation 1-2: GH-2000 Biomarker score for females [47]

$$-8.459 + 2.454 \times \ln(P\text{—} III - NP) + 2.195 \times \ln(IGF - I) - 73.666/age$$

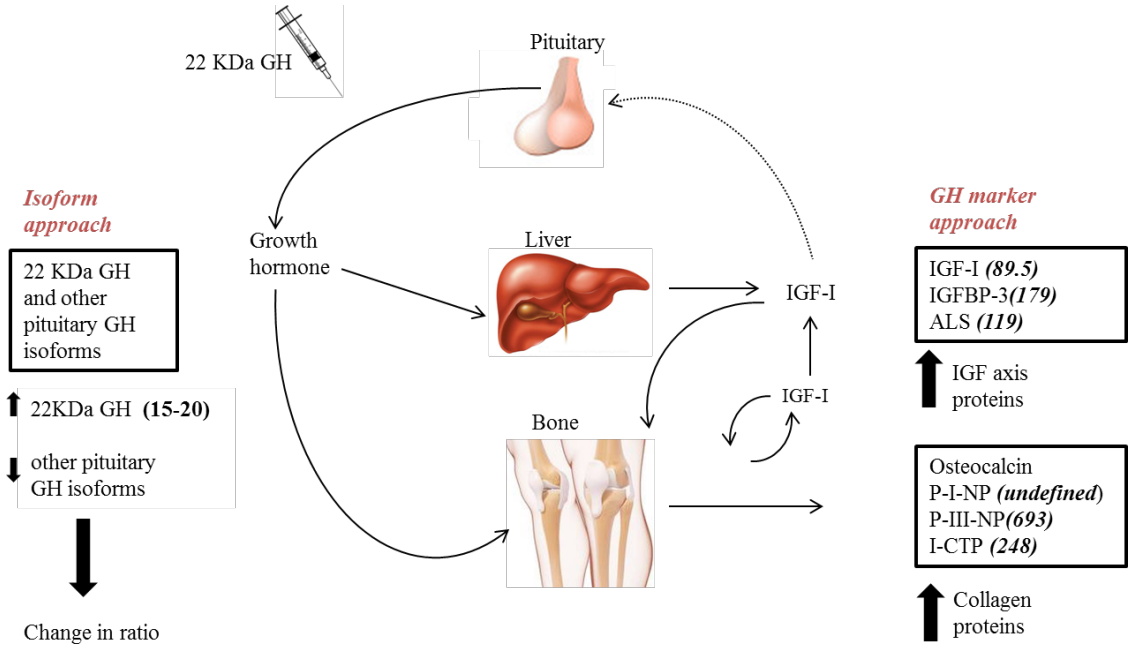


Figure 1.2-1: GH direct (isoform) and indirect (marker) approaches to measurement. Endogenous hGH, secreted by the anterior pituitary, circulates as heterogenous mixture of different isoforms. In the direct measurement approach, differences in the change of hGH ratios caused by the administration of rhGH (containing the 22 kDa isoform) are measured. Whereas in the indirect biomarker approach, measurement of the change in hGH responsive IGF axis proteins and collagen markers are measured. The half-lives for hGH (min) and its responsive biomarkers (h) are shown in **bold**. (Image modified from [49] with permission).

IGF-I and P-III-NP were successfully used to detect hGH administration by the Drug Control Centre (King's College London), whilst conducting the anti-doping tests for the London 2012 Paralympic Games. Two Russian power lifters were caught cheating with hGH based on the anomalous findings of the biomarker test, although they passed the isofom test. Both athletes admitted to doping and were disqualified from the Games. These results demonstrate the advantages of the extended window of detection provided by this approach, which allows for the detection of hGH-related doping days after cessation of drug administration.

1.3 **IGF-I**

IGF-I is a 7.6 kDa protein that is synthesized in the liver and consists of 70 amino acids, six of which are cysteine that form three intra-molecular disulphide bridges [50]. Clinically, serum IGF-I concentration is significant in the diagnosis and treatment of GH disorders. A number of assays exist for IGF-I, which include radioimmunoassays (RIA), enzyme-linked immunosorbance (ELISA), immunochemiluminescence, immuno-functional, and kinase receptor activation [51]. These assays use different antibodies, sample extraction methods (to remove high-affinity binding proteins) and approaches to calibration that have led to a lack in the harmonisation of inter-laboratory results [52, 53]. The typical concentration of serum IGF-I has been shown to be within the magnitude of several hundred nanograms per millilitre (ng/mL) [52], however, ~98 % of the protein is bound to binding protein in serum. Therefore enrichment or serum depletion is necessary before quantification, in order to achieve adequate sensitivity. Methods for the quantification of intact [54-56] and trypsin digested IGF-I [57] by mass spectrometry (MS) have been developed; these complement the existing immunoassays and provide better inter-laboratory agreement of measurements [57].

1.4 **P-III-NP**

Type III collagen is the second most prominent protein present in most normal tissues and plays a major role during growth and tissue repair after injury [58, 59]. Like all collagens, type III is made up of three α chains containing a predominant Gly-X-Y sequence repeat (X and Y represent any amino acid). Several cells possess the ability to synthesise collagen III, however under normal conditions this is done by the fibroblasts using the code of the COL3A gene. Type III collagen is expressed as a procollagen precursor, which in addition to collagen III has two extension propeptides at the amino- and carboxyl-terminal of the protein, referred to as P-III-NP and P-III-CP, respectively [60].

Procollagen synthesis involves several specific enzymes that coil the individual proα1-chains to form intra-chain bonds to result in a helix. During, or after secretion from the collagen producing cell the carboxyl-terminal is lost, leaving pN-collagen. In the extra-cellular space, these collagen molecules form covalent intra- and inter-molecular cross-links to produce collagen fibrils. During the formation of these fibrils P-III-NP is sequentially cleaved from the surface to establish stability in the intra-molecular cross-links and allow for attachment of a new pN-collagen, as shown in Figure 1-3 below [58]. Hence, pN-collagen remains on the surface of the fibril and is only released into circulation during the growth or degradation of the type III collagen fibril. It has been suggested that both P-III-NP (also in a truncated Col 1 form) and P-III-CP regulate the biosynthesis of collagen III using negative feedback mechanisms [61, 62].

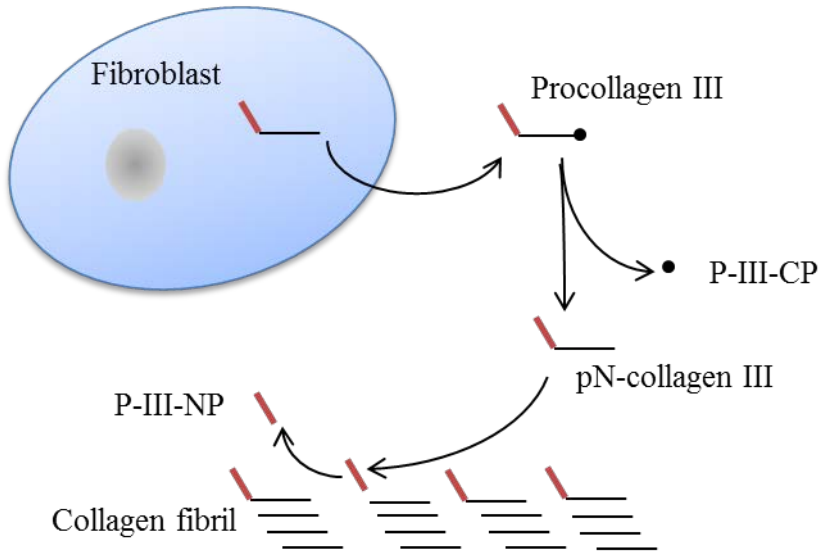


Figure 1.4-1: Synthesis of type III collagen. Secretion of procollagen from the cell leads to the removal of P-III-CP, leaving type III pN-collagen (P-III-NP and collagen). As pN-collagen is attached to the collagen fibril, P-III-NP from the surface is released into the blood stream. (Modified from [59]).

Like collagen, P-III-NP consists of three identical pro α1-chains that are covalently linked by inter-molecular disulphide bonds in the carboxyl-terminal region to form a 42 kDa protein [63].Isolated from pN-collagen with bacterial collagenase, P-III-NP

separates to form three distinct domains, Col 1, 2 and 3, that show different mobility (based on molecular weight (MW)) and can be separated by gel electophoresis [64].

The heaviest fraction, Col 1, can be found at the amino-terminal of the propeptide, and contains three identical peptide chains with 85 amino acids each. This domain is rich in acidic amino acids and cysteine, and is characterised by five intra-molecular covalent disulphide bonds in each chain, which results in the folding of this domain into several loops to form a globular domain [64-66]. At the carboxyl-terminal of the Col 1 is Col 3, the tropocollagen or "collagen-like" domain. Similar to collagen, Col 3 has a tri-peptidic Gly-X-Y repeat, with a high Gly and Pro content that allows the three pro α1-chains in this region to twist into a left-handed super helix [64, 65]. The remainder of P-III-NP is the Col 2 domain, which protrudes from the tropocollagen Col 3. Unlike the prior domain, the amino acids in this region are fully extended; a conformation that facilitates the formation of inter-molecular disulphide bonds between the three chains, which are important for the cross-linking of the Col 3 triple helical domain [66]. In the bacterial collagenase fraction, Col 2 domain overlaps with collagen, to contain seventeen non-P-III-NP amino acid residues from the amino-terminal of type III collagen [65, 67]. Physiologically, the telopeptide region is cleaved during protein maturation by proteolysis at P-130 by type III procollagen N-proteinase to leave three linked 130 aa Col 1-3 pro α1-chains known as P-III-NP [65].

1.4.1 ***Protein sequence***

Prior to the use of mass spectrometry (MS), protein aa sequences were obtained by Edman degradation of peptide fragments (ideally 30 aa in length). In this sequencing approach the amino-terminus residue is sequentially labelled and cleaved from the peptide by reacting with phenyl isothiocyanate, as shown in Figure 1-4 below. As residues are removed their phenylthiohydantoin derivatives are analysed by chromatography or electrophoresis, for identification relative to known standards. However, if the amino-terminal residue is chemically modified (e.g. by acetylation or pyroglutamation) the sequencing of the peptide is blocked, as the five-membered ring

intermediate cannot be formed. For P-III-NP, after removal of the first amino acid residue (Q), Edman degradation was used to sequence the extended Col 1-3 of bovine P-III-NP [65, 67]; that with aa substitutions inferred from the cDNA of the procollagen III α1 human [68] give the aa sequence shown in Figure 1-5.

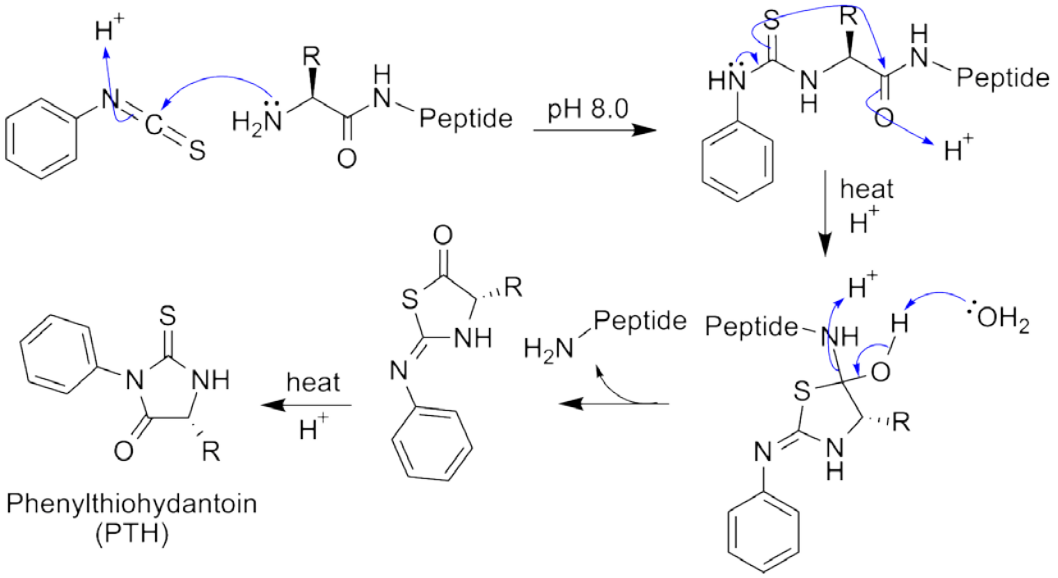
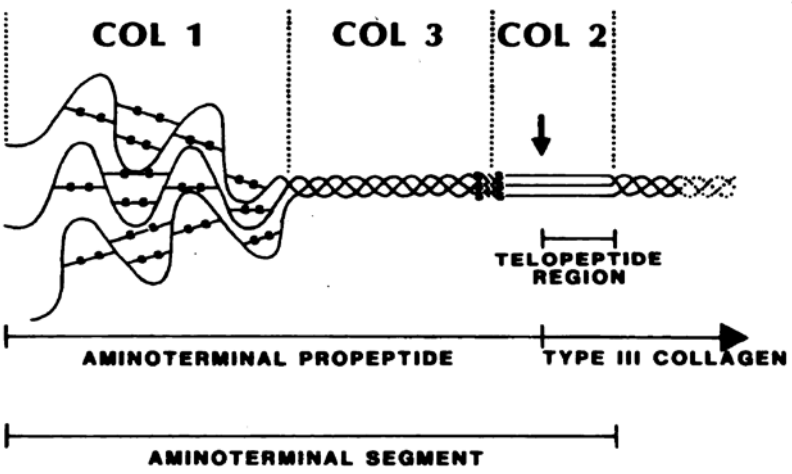


Figure 1.4-2: Showing the Edman degradation mechanism where phenyl isothiocyanate reacts with an uncharged amino group at pH 8 to form a five-membered cyclical phenylthiocarbamoyl derivative that is later cleaved from the peptide under acidic conditions as a thiazoline derivative. The thiazolinone is then treated with acid to form a phenylthiohydantoin (PTH)-amino acid derivative that can be identified by chromatography or electrophoresis. (Mechanism taken from [69]).



Human P-III-NP sequence (UniProt accession no. P02461):

[1]QQEAVEGGCSHLGQSYADRDVWKPEPCQICVCDSGSVLCDDIICDDQELDCPNPEIPFGECCAVCPQPPAPTRPPN
GQGPQGPK[86]**GDPGPPGIPGRNGDPGIPGQPGSPGSPGPP**GICESCPTGPQNYSP[131]*QYEAYDVKSGVAGGGIA*[147]

Figure 1.4-3: Schematic illustration of P-III-NP showing the amino acid (aa) sequence of one of the three identical pro α1-chains. The bars represent disulphide bonds in the Col 1 and Col 2 domains. The Col 3 domain is highlighted in **bold** to show the dominant Gly-X-Pro repeat, where X represents any aa. The *italicised* residues (at the carboxyl-terminus)show the extended Col 2 domain (17 additional aa), retrieved after cleavage from procollagen with bacterial collagenase. (Image modified from [70] with permission).

1.4.2 *Circulating P-III-NP*

It has been suggested that the main antigenic site for P-III-NP is located between aa 20 and 83 of the Col 1 domain [71, 72], although some antigenic properties have been associated with Col 3 or the amino-terminal end of Col 2 [58]. For antigenic activity to occur, the tertiary conformation of the Col 1 domain or the intact P-III-NP protein is necessary [70]. Using the available radioactive immunoassays (RIA), several studies [58, 70, 72-75] have separated (from different body fluids) four antigenic P-III-NP fractions (assigned A-D) by gelfiltration, these are shown in Figure 1-6 below.

In serum (and other body fluids except bronchiolar lavage fluid), main antigenic response was obtained from fraction D, which has a smaller molecular weight than intact P-III-NP, and correlates with the mass of Col 1 [70, 74]. Fraction C corresponds to intact P-III-NP, which interestingly was found in all body fluids except urine [59]. Fraction B (found is serum, wound and synovial fluid) has a MW twice that of P-III-NP and is believed to be a dimer, formed by the covalent cross-linking of propeptide

molecules [58, 76]. Fraction A, corresponding to a higher MW than B, is thought to be complexes of P-III-NP and plasma proteins suspected to have formed during circulation, as it is not found in wound or synovial fluid.

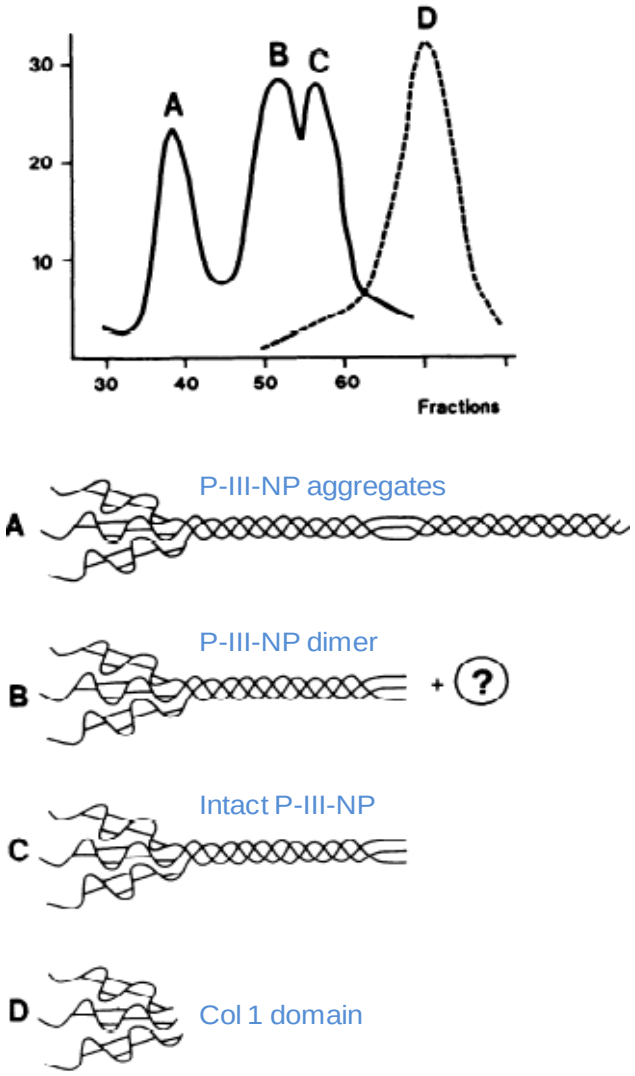


Figure 1.4-4: Elution profile according to Jensen et al.[77]obtained from the antigenic fractions of serum P-III-NP after chromatographic separation using a Sephacryl S-300 column. The solid line show profiles obtained from P-III-NP RIA and the dotted line represent that obtained using a P-III-NP fab RIA kit. The suspected P-III-NP structure for each obtained fraction is represented. (Image taken from [77] with permission).

1.5 **Existing methods of analysing P-III-NP in biological samples**

Like IGF-I, P-III-NP is clinically significant in that it is used to monitor chronic active hepatitis, liver fibrosis and cirrhosis [78-81]. As mentioned earlier, P-III-NP is monitored as a biomarker of GH to aid in detecting administration by athletes in sports doping

tests [82, 83]. For both clinical and anti-doping purposes, quantification of P-III-NP is only achievable by immunoassay approaches. For decades, P-III-NP measurements have been reliant on the use of two radioactive immunoassays, P-III-NP RIA-gnost® (CisBio Bioassays, Codolet, France) and UniQ® P-III-NP RIA (Orion Diagnostica, Espoo, Finland). However, more recently two ELISA methods, Advia Centaur® immunoassay platform (Siemens Healthcare Diagnostics Inc., New York, USA) and P-3NP-EL-US (CisBio Bioassays, Codolet, France) have become available. For all these assays P-III-NP concentrations are expressed as ng/mL, except for the RIA-gnost, which measures in units/mL.

1.5.1 ***P-III-NP antibodies***

Traditionally antibody-based assays (immunoassays), such as ELISA and RIA, were used for the targeted quantification of proteins. Although allowing for high-throughput and sensitive analysis of proteins, the accuracy and precision of these measurements heavily rely on the specificity of the antibodies used. The development of highly specific antibodies can be expensive, time-consuming and resource intensive [84]. Irrespective of the quality of the antibody some cross-reactivity for closely related isoforms or interferents exist. The binding of these non-targeted molecules to the assay antibodies can produce erroneous results.

For immunoassays antibodies can be either polyclonal or monoclonal. Polyclonal antibodies are less specific as they recognise multiple epitopes on the target antigen, whereas monoclonal antibodies recognise only one epitope. Polyclonal antibodies are useful in maximising sensitivity for low abundant antigens, however they are more prone to erroneous measurements. Hence for quantitative immunoassays monoclonal antibodies are preferentially employed, as they show a higher degree of affinity and specificity towards their target analyte.

The antibodies used in the P-III-NP assays are all monoclonal and are raised against purified material from either human or bovine sources [85]. However, the binding epitope of these antibodies are not well characterised and cross-reactivity for different

P-III-NP species is observed [59]. In the absence of authentic human P-III-NP (*h*P-III-NP) material, all the immunoassays use bovine P-III-NP (*b*P-III-NP), which shares 97 % sequence homology with *h*P-III-NP, as the assay quality control sample [71, 85, 86].

1.5.2 ***Limitations***

For the RIAs, the use of radioactive material as the signal-generating label associate drawbacks to methods. Radioactive materials are health hazards, thus special licensing is needed for handling such material and handlers need to be properly trained. The storage of kits are limited by the short half-lives of the isotopes used, including iodine-125 (59.49 days ± 0.13) that are used in P-III-NP RIA kits [87]. Also specific instruments are required for radioactive counting.

For the Advia Centaur® immunoassay, a specific platform is needed. In order to measure P-III-NP samples, the Advia Centaur® is reliant on instrument calibration using the manufacturer provided low and high quality controls, for which the components and their concentrations are undisclosed. As the most recent assay to be released, the P3NP-EL-US is still awaiting validation by anti-doping laboratories, before it can be considered for use in sports testing. All the other assays listed above have been validated and approved by WADA for determining P-III-NP concentrations, to be used in the detection of GH administration. A lack of commutability exists with the P-III-NP measurements produced the available assays, where measurements for the same sample using each assay gives a different result. As a consequence of this, different GH-2000 score calculations are employed for the different P-III-NP and IGF-I assay pairs used to measure analyte concentration.

The quality of immunoassay measurements is heavily reliant on the antibodies used. Due to the difficulty and cost associated with antibody manufacture, immunoassay kits are prone to change or withdrawal from the market. If and when these changes occur, validation and re-assignment of GH-2000 score decision limits become necessary. To overcome these and other limitations, analytical methods based on the mass

spectrometric measurement of P-III-NP are desirable [3]. However, as a result of its size (42 kDa) and the sub-nanomolar serum concentration (25-125 pM), the development of such methods for P-III-NP is difficult, especially with the lack of reference material.

1.6 **Protein mass spectrometry**

Significant advances to mass spectrometry (MS) over the last few decades have resulted in the use of MS-based assays for protein quantification as an alternative to immunoassay approaches. High specificity and selectivity are associated with MS-based assays, especially where unique selected reaction monitoring (SRM) methods are used to target the specified protein. The proteins targeted in these bio-analyses are commonly found in complex biological matrices such as serum. Prior separation of these analytes from other matrix proteins is necessary in order to obtain sensitive and reproducible results from MS methods. This can be achieved with appropriate sample preparation and chromatographic separation techniques, which are usually developed in parallel to MS methods so as to produce complete bioassays.

1.6.1 ***Sample preparation***

A wide dynamic range of protein concentrations exist in biological matrices such as serum and blood. The complexity of these matrices negatively affect the detection (and quantification) of low-abundant proteins (LAPs) such as P-III-NP by MS. This is due to the presence of high-abundant proteins (HAPs), like albumin and IgG, which mask the detection of the LAPs by causing ion suppression in the mass spectrometer. Thus simplification of the sample matrix is necessary. In proteomics, this is achieved through the fractionation of the biological sample by depletion or enrichment approaches.

Due to the wide variance of physical and chemical characteristics associated with proteins (as a result of different aa compositions), there is no model for sample preparation for this class of compound. Successful methods that have been used to deplete HAPs in proteomic studies include affinity chromatography, ultrafiltration,

dialysis and protein precipitation [88-91]. These techniques vary in mechanism and efficiency, thus they should be individually assessed for applicability to each intended bioassay. Although these techniques are targeted towards the depletion of HAPs, losses of LAPs may occur if they are bound to HAPs as protein complexes. Disruption of the binding within these complexes can be achieved by changing the sample pH through the addition of organic acids (e.g. acetic, formic and trifluoroacetic acid), mild bases (e.g. ammonium hydroxide) or denaturing agents (e.g. urea).

As a more targeted and selective approach to sample preparation, the enrichment of LAPs is recommended. This approach utilises unique properties of target proteins such as their aa sequence, post-translational modification (PTMs) (e.g. glycosylation) or antibody epitope to capture and separate them from the sample matrix [92-94]. For antibody-based capture the availability of specific antibodies is a limiting factor of this application. However, unlike immunoassays, polyclonal antibodies are appropriate for quantification methods as antibodies are only used as capture tools and are not involved in detection.

In addition to separating proteins from matrix, peptide handling and storage need to be considered. Non-specific adsorption loss to surfaces such as pipette tips, sample vessels and instrument tubing is commonly observed in protein analysis, especially at low concentrations. It is thought that these losses can be minimised with changing the protein (or peptide) environment through the adjustment of solvent composition, pH and temperature [95]. However, some level of caution should be exercised when altering the protein environment so that its chemical stability (e.g. to oxidation, reduction, and deamidation) and physical stability (e.g. to denaturation and aggregation) is maintained. As an alternative approach to reduce non-specific surface loss, the incorporation of excess amounts of surrogate molecules, which do not interfere with analytical methods and will preferentially bind to non-specific sites in preference to target analytes, is incorporated in the sample preparation protocols.

Although the complexity of samples are reduced using these sample preparation approaches, the probability of complete analyte isolation is small. Residual HAPs and non-targeted LAPs will remain in the fractionated matrix along with the target peptide, however, their concentration ranges should be of the same order of magnitude. For MS based approaches further separation of the sample is required and, to achieve this, liquid chromatography is often used.

1.6.2 ***Liquid chromatography***

For additional separation of the separated sample matrix, which contains some residual HAPs and untargeted LAPs, liquid chromatography (LC) is often used prior to MS for protein analyses [96]. In LC further separation of the sample components is achieved based on their differences in distribution between an immiscible mobile and stationary phase. Reversed-phase LC (RPLC) is commonly used for protein and peptide quantification, where the stationary phase is non-polar and the mobile phase is polar. Using silica based particles as solid support, alkane hydrocarbons (C4 -C18) are tethered to the column, through which the mobile phase is pumped to elute separated analytes. The mobile phase is generally a binary mixture of water and a miscible polar organic solvent (e.g. methanol or acetonitrile). Non-polar analytes will interact with the stationary phase and be retained, however the degree of interaction of each analyte is dependent on its hydrophobicity. By manipulating the composition of the mobile phase (i.e. polar: non-polar ratio), elution of analytes from the column stationary phase can be controlled to effectively separate the sample mixture. Alternative phases such as normal-phase or hydrophobic interaction LC (HILIC) have been applied to protein analysis [95, 97].

1.6.3 ***Mass spectrometry***

Prior to 1990, the use of MS was limited to the analysis of small volatile molecules (< 1000 Da) for which gas-phase charged ions (measured as mass to charge (*m/z*) in the MS) could be formed using electron ionisation (EI), a hard ionisation technique. The high energy transferred in EI makes it inappropriate for the analysis of large molecules,

like proteins, nucleotides and carbohydrates, as they would be completely destroyed or degraded during ionisation. However with the invention of soft ionisation techniques (i.e. electrospray ionisation (ESI) and matrix assisted laser desorption ionisation (MALDI)) and the potential coupling of MS to LC, the analysis of these large molecules by MS was made possible. Interfaced with LC, ESI is the most appropriate ionisation mode. This is because ESI forms gas phase ions from solvated sample (described below in section 1.6.3.1), whereas MALDI does this using a solid state (section 1.6.3.2).

1.6.3.1 Electrospray ionisation

In ESI, solutions eluted from LC are infused under atmospheric pressure into the ionisation source through a capillary, where they are nebulised with a high electric voltage (3-4 kV) to form an electrospray containing highly charged droplets. Once formed these droplets are electrically driven towards the entrance of the mass spectrometer with the aid of warm neutral gas (usually nitrogen). Two mechanisms for the formation of gas phase ions from these charged droplets have been proposed; these are the ion evaporation and charge residual ionisation models [98, 99]. Using the ion evaporation model, LC solvent is evaporated and the droplets break down to continually reduce in size. Eventually the repulsive force of the ions on the surface of the shrinking droplet becomes high enough to exceed the surface tension of the solvent, thereby producing ions that are in the gas phase [100]. Alternatively gas phase ions are produced using the charge residue model, where the continuous evaporation of the solvent accompanies droplet fragmentation to leave a single ion (completely devoid of solvent) at the end of the process [98]. Figure 1-7 shows the differences of these processes. ESI tends to produce multiply-charged ions from biomolecules such as proteins and peptides; by doing this the theoretical mass limit of the MS is extended. Thus large molecules will appear with a lower *m/z* despite their mass being beyond the range of the MS.

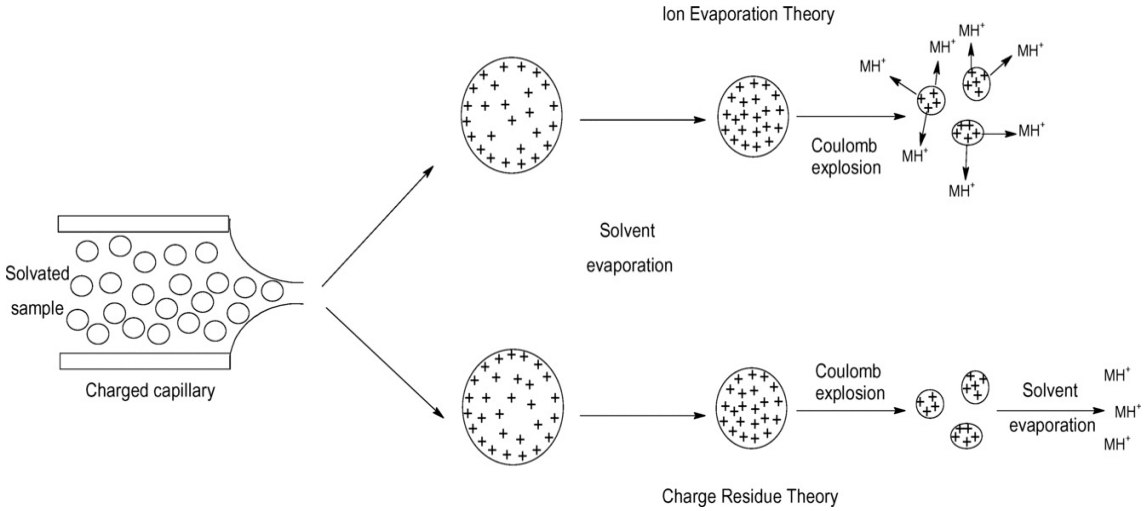


Figure 1.6-1: Schematic of ESI , showing the ion evaporation and charge residual theories for the formation of gas phase ions. Solvated sample is passed through a capillary to form charged droplets, which are reduced to single ions by columbic explosion (ion evaporation model) or solvent evaporation (charge residual model). The scheme represents positively charged ions, however, the same mechanism applies in negative ion mode. (Image taken from [101] with permission).

1.6.3.2 Matrix assisted laser desorption ionisation

In MALDI, the sample is mixed with an appropriate solid matrix and an ultraviolet laser beam (usually a nitrogen laser at 337 nM) is used to transfer energy to the matrix and then the analyte to result in their subsequent ionisation and desorption from the surface. With MALDI, ions are mostly generated as single charged species. The process of ion formation is shown in Figure 1-8. Unlike ESI, the mechanism of ion formation for MALDI is not fully understood, and the matrix used is mainly experimental although it may be influenced by ionisation mode [101, 102].

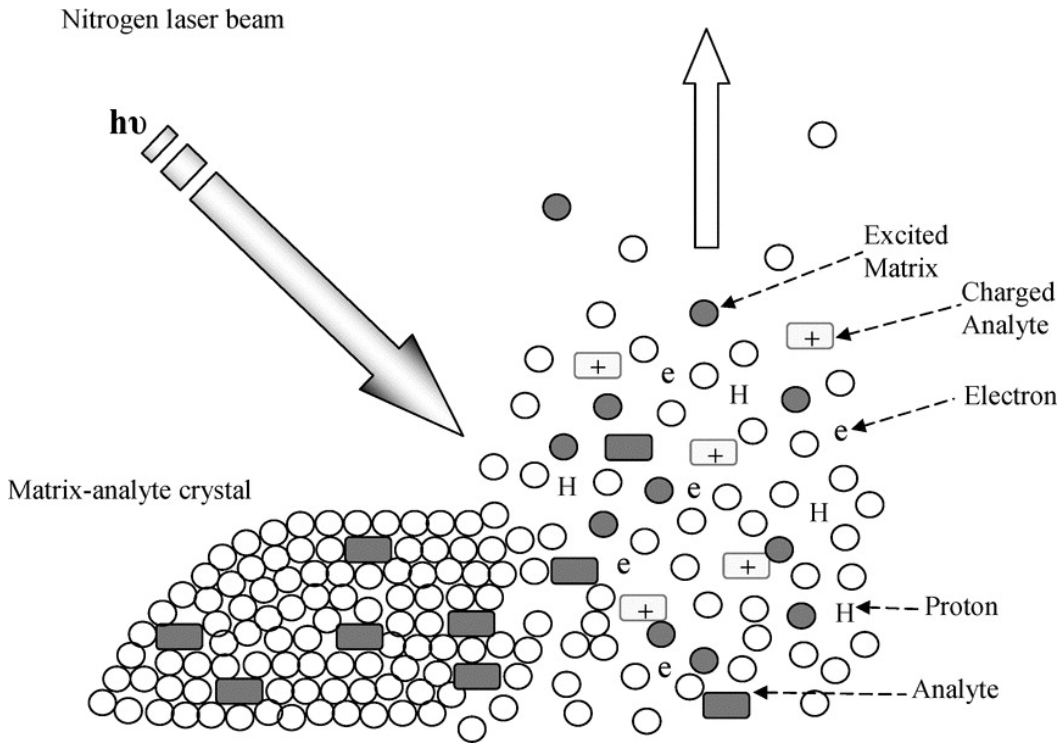


Figure 1.6-2: MALDI ionisation process, where matrix-analyte crystals are excited with a laser beam to result in the ionisation and desorption of the analyte from the surface. (Image taken from [101] with permission).

1.6.4 *Mass analysers*

The four main mass analysers used in proteomics are linear quadrupoles, time of flight (TOF), trapping quadrupoles and the Orbitrap [103]. MS which combine these analysers (e.g. quadrupole-TOF) are also commonly used.

1.6.4.1 Quadrupoles- linear and ion trap

Linear quadrupoles consists of four equally spaced and shaped rods to which a specific radiofrequency (RF) and direct current (DC) is applied. Manipulation of these voltages allows for the selection of ions with a specific *m/z* based on the stability of the trajectory of the ions [104]. Ion trap quadruopoles (QIT) use an oscillating field to store ions of multiple m/z values, based on the RF of a quadrupolar field in a concentric three-dimensional orbital. The 3D QIT has two end cap electrodes that are electrically isolated from either side of a ring electrode. Here ions of all masses are trapped and form different trajectories around the central electrode, where an RF is applied that causes the ions to oscillate with a specific radius based on their m/z. By altering the

amplitude of the RF applied to the electrodes, ion trajectories are eventually destabilised to result in the expulsion of ions from the trap (according to mass) [104].

Quadrupoles have a limited mass range (usually < 4000 Da) and are considered to be low resolution instruments as they have unit mass resolution, however they give good sensitivity owing to their mass filtering capabilities which reduces background noise [105].

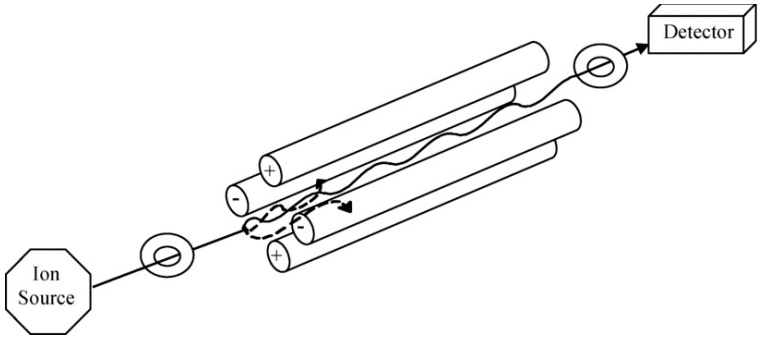


Figure 1.6-3: Schematic of a linear quadrupole mass analyser showing the four parallel electrical rods to which differing direct current and radio-frequencies allow for separation of charged ions based on their trajectory. The solid line show ions with the "right trajectory" which reach the detector, and the dotted line show ions with the "wrong" trajectory which collide with the rods and are ejected. (Image taken from [101] with permission).

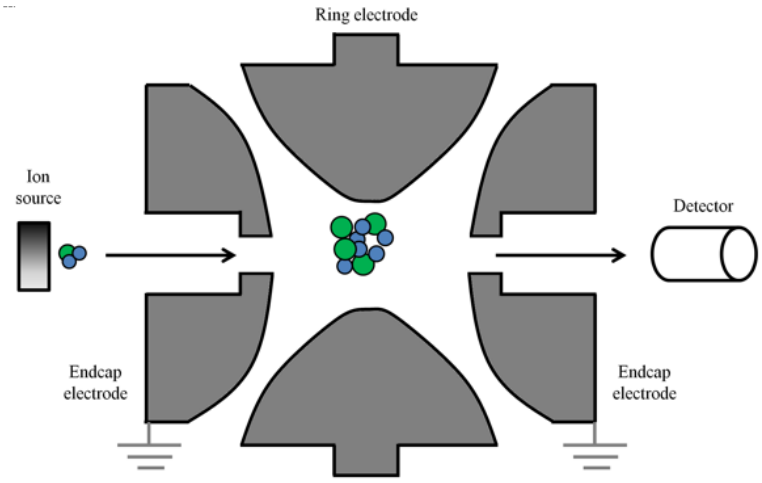


Figure 1.6-4: Schematic of a 3D ion trap mass analyser showing the ring and end-capped electrodes, where alternating and direct potentials are applied to eject the trapped ions (oscillating at different trajectories) according to mass. (Image taken from [106]).

1.6.4.2 Time of flight

Separation of ionised species using ToF relies on differences in the analyte's free flight (associated with mass) while travelling in a field free tube (1-2 m) to reach the detector. With this approach all formed ions will reach the detector, unlike quadrupoles that filters ions. The mass range of a ToF is theoretically limitless, although the resolving power is dependent on tube linearity and length [107]. However, the use of a reflectron-type mass analyser with an electrostatic ion mirror (shown in Figure 1-11) can extend the flight path of ions with higher kinetic energy, which penetrate deeper into the mirror, and are gradually expelled to further separate ions and improve the resolution of the ToF spectrum.

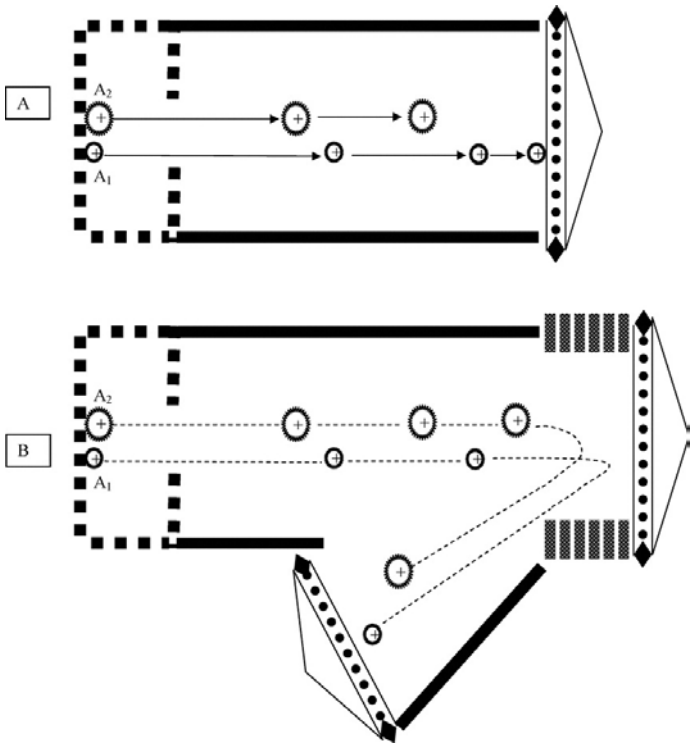


Figure 1.6-5: Schematic of flight path of charged ions in a conventional ToF (A) and reflectron-type ToF (B) mass analysers. (Image taken from [101] with permission).

1.6.4.3 Orbitrap

Orbitraps have two end cap electrodes enclosing a central electrode, which unlike quadrupoles, uses a static electrostatic field to sustain ion trapping following a specialized dynamic injection pulse. As the ions enter the Orbitrap they oscillate around

the central electrode. Here each ion has a characteristic frequency that is converted to a *m/z* by Fourier transformation to produce a spectrum with accurate mass measurements (accuracy < 5 ppm) and high resolution (> 140,000 FWHM) [86]. Like ToFs, all the ions that enter the MS will reach the detector; thus this instrument is only capable of full scan experiments unless coupled with a quadrupole mass analyser.

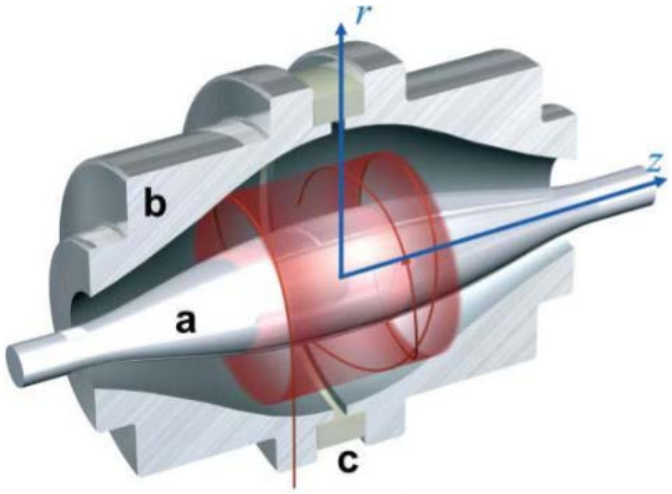


Figure 1.6-6: Cut-away model of an Orbitrap mass analyser, showing the oscillation path of ions around the central electrode(a) and outer electrode (b) (which is split in half by an insulating ceramic ring, (c)). (Image taken from [108] with permission).

1.6.4.4 Ion cyclotron resonance

Ion cyclotron resonance (ICR) is related to the movements of ions in a magnetic field. Where ions in a cyclotron at the same frequency as the electromagnetic wave adsorb energy to increase the radius of its trajectory to 1 cm, at which point they strike a perpendicular to create a measurable "image current" [109]. The frequencies appear as a complicated waveform consisting of an overlay of the individual frequencies and amplitudes that can be converted by Fourier transformation as an infinite series of sine and cosine functions that are expressed as a mass spectrum [104].

1.6.5 ***Tandem MS***

So far the analysers as described are only capable of single MS analysis (e.g. full scan or single ion monitoring (SIM)) that can be used for molecular ion determination. However if quadrupoles are placed in tandem with each analyser (i.e. quadrupole-quadrupole (QqQ), quadrupole-ToF (Q-ToF) and quadrupole-Orbitrap), MS/MS experiments for a target precursor can be conducted. In this set up the quadrupole is used to isolate target ions that can be fragmented in a collision cell before entering the second analyser for *m/z* separation (schematic shown in Figure 1-13 below).

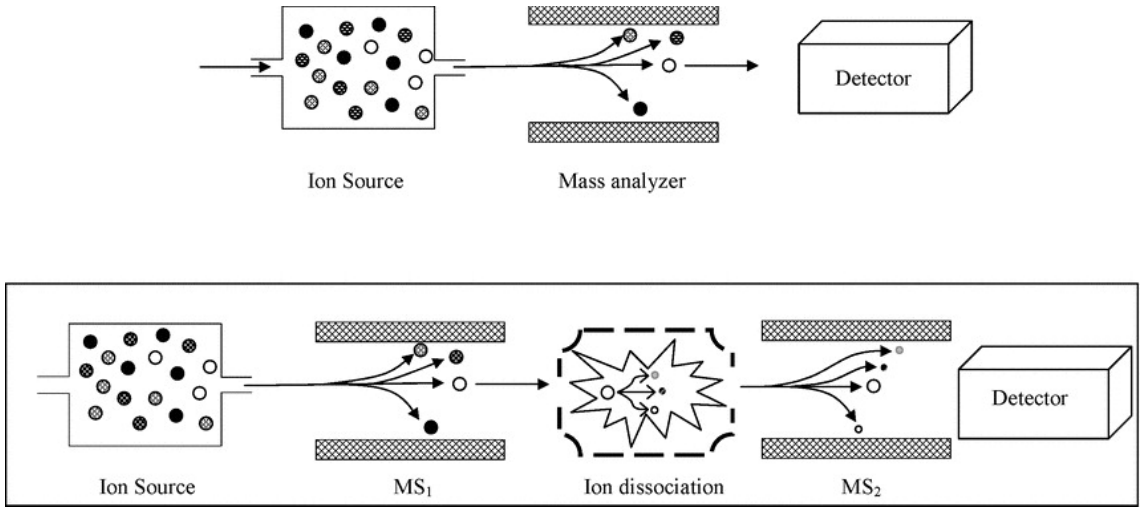


Figure 1.6-7: Comparison of single (top) and tandem (bottom) MS analysers. In single MS , ions that enter the MS are separated in the analyser before reaching the detector. In tandem MS, specific ions can be selected in MS_1 and fragmented, fragments are then separated in a second analyser (MS_2) and detected. (Image taken from [101] with permission).

For the purpose of this project, peptide characterisation was performed using a Q-Exactive® (quadrupole coupled to an Orbitrap) system, which produces untargeted high resolution measurements (to 4 decimal places) from full and product ion scan acquisitions. In full scan, samples are analysed directly by the Orbitrap using HRAMS (high resolution accurate MS). For product ion scanning, analytes of a specific mass range are filtered by the quadrupole MS into a collision cell, where fragmentation occurs prior to HRAM analysis in the Orbitrap. However in developing targeted quantitative methods, triple quadrupole (QqQ) instruments are employed to optimise

the sensitivity of analysis by increasing signal to noise ratios through the mass filtering capabilities provided in SIM. Here multiple unique fragment reactions are monitored in selected reaction monitoring (SRM) mode to increase the selectivity of MS analysis.

1.6.5.1 Protein fragmentation

Soft ionisation of peptides by ESI limit in-source fragmentation and produce $[M+xH]^{x+}$ (where x= 1, 2, 3...) in positive mode. In ESI where mild acidic conditions are used for protonation, multiply charged ions can be produced from proteins, which have multiple protonation sites. Higher charge states (i.e. $x \geq 2$) are common in peptides ≥ 15 aa, or that contain multiple basic residues (Arg, Lys and His). Under MS/MS conditions protonated peptides fragment at Cα-C, C-N or N-Cα by rearrangement-type reactions or direct bond cleavage to produce 6 typical fragment ions [110]; for which their standardised Roepstroff-Fohlmann-Biemann [111, 112] nomenclature is presented in Figure 1-14 below.

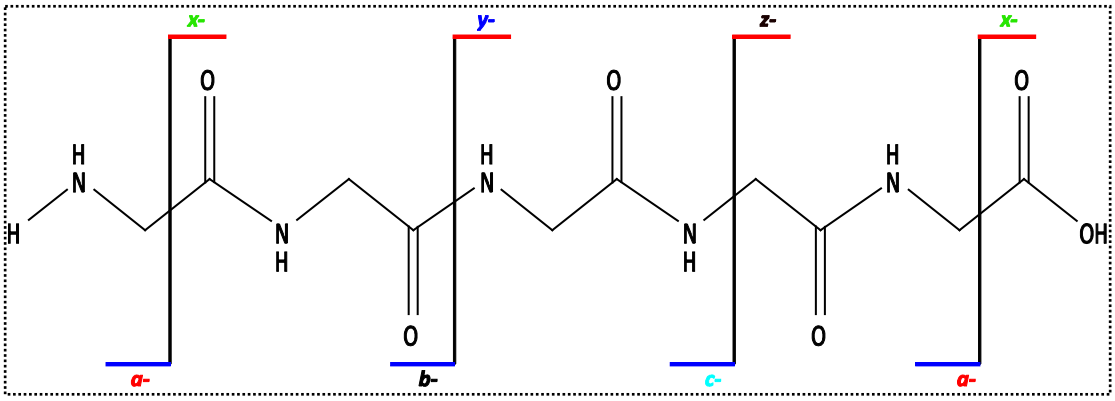


Figure 1.6-8: Nomenclature of the common peptide fragment patterns observed in tandem mass spectrometry [111].

The most comprehensive model to describe the peptide dissociation by re-arrangement is the "mobile proton" model where upon excitation added proton(s) migrate around the peptide to various sites prior to the cleavage of weakened and thermodynamically less stable peptide bonds [113]. Protons that are sequestered by side chains of basic residues are unable to direct fragmentation by this model. Lower energy is required for

re-arrangement type reactions than is needed for direct bond cleavage, making the latter mode of fragmentation uncommon in ESI-MS [110].

1.6.6 ***Quantification of proteins using MS***

With this application of MS to protein analysis accurate protein mass measurements are now obtainable that has aided the accurate determination of protein (and peptide) sequences and facilitated the identification of post-translational modifications (PTMs). However for MS, proteins need to be within the upper *m/z* range of instruments (currently around 4000-5000 *m/z*), or only require a small amount of multiple charging to achieve a detectable *m/z* of reasonable abundance. Therefore for large proteins, smaller peptides can be obtained by protein digestion (incorporated in the sample preparation step of analysis), which are analysed in place of the intact protein to reduce the spread of the MS signal seen with multiple charging and improve the signal intensity. When taking this approach however, surrogate peptides need to be selective and specific to their precursor protein to ensure MS measurements are of the target protein and not a contaminant.

For quantitative analyses internal standard(s) (ISTD) should be incorporated in the bioassay. The ideal ISTD should have similar physical and chemical characteristics to the target analyte, in order to control any assay variability associated with sample preparation, LC and ionisation efficiencies (associated with matrix effects or drift in ion current). In most MS assays, isotopic analogues of the target analyte are used (most likely containing ^{13}C and/or ^{15}N), as these compounds share chemical properties with the analyte but have different masses, so that they can be separated in the MS analyser to produce separate MS signals [114]. A mass difference of at least 3-4 Da is preferable to minimise errors associated with isotopic overlap [95]. Where digest peptides are targeted in the MS for quantification, isotope labelled peptides can be used in place of the isotope labelled protein standard. Although the peptide has similar chemical and physical properties in LC and MS, in sample preparation where the intact protein is the analyte, these standards cannot compensate for variance.

1.6.7 ***Protein and peptide standards for P-III-NP***

As a low abundant protein whose expression is age dependent, natural sources of *h*P-III-NP are limited. By recombinant approaches, procollagen III has been produced for application to tissue engineering in the medical field [115-117]. It is important to ensure the occurrence of PTM events (e.g. specific Cys-Cys disulphide bonds and Pro hydroxylation), which affect the cross-linking of the protein; these processes are dependent on the age and physiology of the tissue used for maturation during recombination [118]. On the basis of yield (20 % increase) for the target type III collagen, the recombinant production of procollagen III without P-III-NP has predominated in this field [118]. Hence, although possible to produce by recombinant technologies a source to r*h*P-III-NP is not available.

Similar to the *in vivo* production of proteins, peptides and proteins can be chemically synthesised through the gradual elongation of the aa units in its sequence via coupling reactions. Like the recombinant production, PTM and cross-linking is important to protein synthesis and need to be incorporated into the synthesis process. This complicates synthesis, and potentially produces a combination of products of varying and irreproducible yield. Peptide synthesis is a simpler process, where although PTMs are essential for association with the endogenous protein, cross-links are not as essential, especially where they are removed before LC-MS. Synthetic synthesis of peptides can be achieved by solution or solid phase mechanisms, where the latter is practical for research applications based on the quantity of material used [119].

1.6.7.1 Peptide standards - synthesis

For solid phase synthesis, peptides are synthesised from the carboxyl- to amino-terminus of the peptide. To start the peptide synthesis the first aa (carboxyl-terminus aa in the sequence) is coupled onto an activated linker on the surface of the solid-phase resin. This is an important step in synthesis, in that the efficiency of this coupling defines the yield and scale of SPPS. Once attached the protecting molecule from the α-amino group is selectively removed to provide a free amino group, able to react with

the chemically activated α-carboxyl group of the subsequent aa for coupling. Peptides are elongated with the repetitive coupling and de-protection of aa residues until the sequence is complete. After this the synthesised peptide is removed from the resin along with any additional reactive side chain to yield the desired product. The generalised scheme of SPPS is shown in Figure 1-15 below.

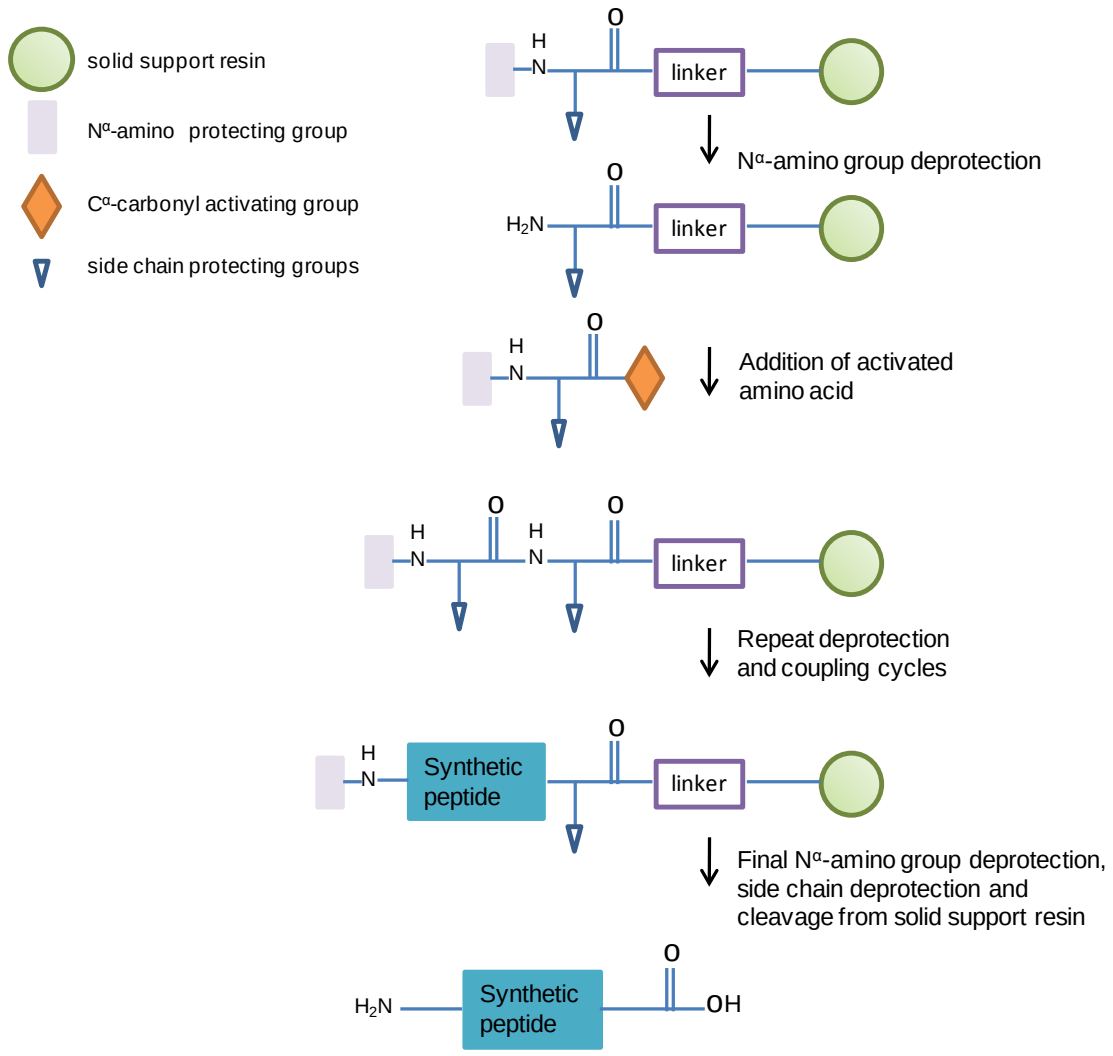


Figure 1.6-9: Generalised approach to solid phase peptide synthesis, where peptides are synthesised on a solid support resin from their carboxyl- to amino- terminus. During their synthesis peptides are elongated through by the sequential addition of aa using a repetitive coupling and de-protection mechanism. To dictate the location of amide bonds, aa residues with reversible-protected side groups (e.g. Fmoc, Boc and Trt) are used.

1.7 **Aims and objectives of this PhD**

The overall aim of this PhD is to demonstrate that LC-MS is capable of quantification of serum *h*P-III-NP using a digest MS approach and to complement the existing immunoassays, as a part of the evidence towards determining hGH administration in sport doping tests.

1.7.1 ***Objective 1: To evaluate the yield of digest peptides from P-III-NP***

In the absence of international reference material for P-III-NP, partial characterisation of the protein is undertaken by enzymatic digestion using the aa sequence (obtained from genomic databases) and a sample of *b*P-III-NP (sharing 97 % sequence homology with *h*P-III-NP). Following evaluation for selectivity and MS sensitivity, surrogate peptides were selected for developing target LC-MS methods for *h*P-III-NP. These methods are covered in Chapter 2.

1.7.2 ***Objective 2: To develop sensitive and selective LC-MS methods targeting P-III-NP***

Using the selected peptides, targeted LC-MS methods were developed to achieve the optimum sensitivity for P-III-NP, so as to enable analysis at the reported serum level (1-5 ng/mL, 25-125 pM). To achieve this, conventional, micro- and nano-flow technologies are investigated in Chapter 3.

1.7.3 ***Objective 3: To develop a suitable sample preparation procedure for serum P-III-NP analysis***

Present within complex matrices (i.e. blood and serum) at very low concentrations the direct analysis of P-III-NP by MS is not possible without sample preparation. In Chapter 4 is a description of some of the common techniques used in proteomics that have been investigated to assess their efficiency and applicability for separating P-III-NP from serum for MS analysis.

1.7.4 ***Objective 4: To develop an approach to obtaining peptide standards for the quantification of P-III-NP***

Once suitable methods for P-III-NP analysis in serum have been developed, appropriate standards for the quantification of P-III-NP need to be obtained. Initial synthesis of labelled and unlabelled peptide standards are presented in Chapter 5. With successful development of these standards, quantification of serum *h*P-III-NP by LC-MS should be achieved.

Chapter 2 *: In silico and in vivo digestion of P-III-NP*

2.1 **Overview**

This chapter details the *in silico* and *in vivo* digestions of procollagen III amino terminal propeptide (P-III-NP) to predict peptides to target for quantitative mass spectrometry (MS). P-III-NP was *in silico* digested with five endoproteases (trypsin, Asp-N, Glu-C, chymotrypsin and Lys-C) to identify the best digest approach to support development of a selective quantitative LC-MS method. Bovine P-III-NP was digested with trypsin prior to partial *de novo* peptide sequencing using liquid chromatography tandem mass spectrometry (LC-MS). Deviations from the peptide masses determined by *in silico* digests were observed in experimental products, due to the presence of post translational modifications (PTM) in some amino acid (aa) residues. The uniqueness, stability and MS response of the digested peptides generated by both *in silico* and *in vivo* methods were assessed to target surrogate peptides for possible quantification of P-III-NP by LC-MS.

2.2 **Introduction**

Mass spectrometry (MS) has become the bioanalytical method of choice for sensitivity and selectivity [97]; with recent advances in MS and chromatography, tandem techniques can now be used to measure small amounts of protein in complex matrices with greater throughput and improved accuracy and precision. Liquid chromatography (LC) is used to separate the many components present in complex biological mixtures (e.g. serum), whilst MS detection allows for the accurate and precise mass measurement of targeted (and untargeted) analytes. Unpredictability of the associated molecular weight of proteins resulting from amino acid (aa) post- translational modification (PTM) across the peptide backbone can complicate MS data interpretation. Hence, protein characterisation (sequencing) is undertaken in the first stages of method development.

2.2.1 ***Procollagen III amino terminal propeptide***

Procollagen III amino terminal propeptide (P-III-NP) is a 42 kDa trimeric protein [64] consisting of three identical pro α1- chains covalently linked by disulphide bonds [59]. Bacterial collagenase digested P-III-NP separates on diethylaminoethyl (DEAE) cellulose to reveal three conformational distinct domains. Carboxyl- to amino-terminus; these are the globular (Col 1), triple helical or "collagen-like" (Col 3) and telopeptide precursor (Col 2) domains [64, 66]. In conjunction with the aa sequence of P-III-NP, the topology of these domains help predict possible P-III-NP post-translation modifications that affect protein mass, structure and stability of the protein.

2.2.1.1 Deamidation

Glutamine (Gln) and asparagine (Asn*)* residues in peptides and proteins are inherently unstable under physiological solvent conditions and are often deamidated to glutamic acid (Glu) or pyroglutamate (*pyro*E) and aspartic acid (Asp) respectively; in a wide variety of proteins [120]. Deamidation of Gln is much slower than Asn; except for amino-terminus Gln, where deamidation occurs rapidly to form a *pyro*E [121].

2.2.1.1.1 Asparagine

Deamidation of Asn is not uncommon in proteins where it is amino-terminal to glycine (Gly) and to a lesser extent to serine (Ser) [122]. This process can take place slowly in intact proteins, however the rate of reaction increases dramatically after proteolytic digestion; where residues are left more exposed [123]. Rates of deamidation are dependent on pH, ionic strength, temperature and solvent conditions as well as the structure of the peptide or protein near the residue [120]. A typical conversion of 70-80 % of Asn-Gly → Asp-Gly and ~10 % for Asn-Ser → Asp-Ser has been observed in the commonly used overnight trypsin digestion of proteins [122].

2.2.1.1.2 Glutamine

*Pyro*E is formed enzymatically in the presence of glutaminyl cyclase, but can also occur spontaneously from amino-terminus Gln or Glu through the deamidation or dehydration of respective residues [124]. The reaction rate for Gln is faster than that for

Glu in both the catalysed and spontaneous reactions [125]. Cyclised *pyro*E makes proteins resistant to aminopeptidase degradation thus slowing physiological turnover. Furthermore, it blocks peptide sequencing by Edman degradation, where phenyl isothiocyanate requires a free primary amino group to react. Hence, it is not surprising that early Edman sequencing of P-III-NP (calf) required prior removal of the first aa with pyroglutamate aminopeptidase, thus, providing evidence for this modification [65]. Amino-terminus pyroE has also been identified in calf type I [126] and sheep type I [127] procollagen peptides.

2.2.1.2 Hydroxylation

Col 3 P-III-NP is commonly referred to as the “collagen-like” domain owing to its similarity with collagen in topology and basic aa structure. Collagen has a basic triple helical conformation consisting of three supercoiled poly-proline (Pro) helices that require a Gly residue at every third position to obtain a predominant Gly-Pro-Pro repeat [128]. Hydroxyproline (*hydroxy*P) is formed by the enzymatic hydroxylation of Pro by prolyl hydroxylase [129], and is stereo and site specific in that it occurs only at Pro residues in the third position to give a Gly-Pro-*hydroxy*P repeat. This Pro hydroxylation adds stability to collagenous regions by enabling covalent bonding between the hydroxyl group and water molecules to form a tropocollagen [128, 130], as shown in Figure 2-2.

2.2.1.3 Glycosylation

Glycosylation, one of the more complex PTMs, is a common covalent modification in eukaryotic cells [131]. Greater than 50 % of human proteins are glycosylated [132] where the attached glycans are either nitrogen (*N*-) or oxygen (*O*-) linked [131].

2.2.1.3.1 Nitrogen linked glycosylation

N-linked glycosylation can occur at the side chain amide nitrogen of Asn where a Asn-X-Ser/Thr sequon exists; X represents any aa residue except Pro. These glycans are added *en bloc* to the protein in the lumen of the endoplasmic reticulum usually as pre-synthesized core units of 14 saccharides [133] often involving GlcNac

(*N*-acetylglucosamine) [134, 135]. Among its many functions, *N*-linked glycosylation helps to stabilise proteins against denaturation and proteolysis, enhance solubility, facilitate orientation of proteins relative to a membrane, confer structural rigidity and regulate protein turnover [134].

2.2.1.3.2 Oxygen linked glycosylation

O-linked glycosylation takes place in the Golgi apparatus where linkage occurs by the transfer of GalNac (*N*-acetylgalactosamine) to Ser/Thr by *N*-acetyl galactosaminyl transferase. After attachment of the first GalNac residue, gradual elongation and modification of the core structure can occur [136]. *O*-linked glycosylation sites are not easy to predict, as sugars are often clustered in short peptide regions rich in Ser and Thr. *O*-linked glycosylation can modulate aggregation, maintain stability of and confer protease and heat resistance to proteins [133].

2.2.2 ***Protein digestion***

Successful characterisation of P-III-NP PTMs will allow for the correct calculation of the protein's molecular weight. In general the MS sensitivity of proteins is less than that for peptides. As most MS instruments have a typical mass cut-off *m/z* 3000, proteins need to be multiply charged for analysis; thus, the peak intensity is shared between these multiple charges to effectively reduce the overall MS sensitivity. Ideal sequencing data is obtained from peptides ≤ 20 aa long [103]. As a low abundant serum protein, MS sensitivity for P-III-NP can be maximised by analysing smaller surrogate peptides obtained from endoprotease protein digestion.

Digestion of proteins with trypsin is a standard step in proteomics. Trypsin owes this position to its high efficiency, cleavage site-specificity [137], and suitability of fragments for MS [138]. The cleavage-site specificity of trypsin is due to its highly conserved catalytic triad (Ser-195, Asp-102, His-57, shown in Figure 2-1 below) and the specificity of the binding pocket for basic residues arginine (Arg) and lysine (Lys) [139]. The frequency of trypsin cleavage in proteins generally yield peptides 7-35 aa long with a mass range of 600-4000 Da [140]. Most tryptic peptides have a carboxyl-terminus

Arg/Lys that enhances positive ionization and leads to favourable fragmentation of peptides potentially dominant in y-ions [138]. Trypsin “missed cleavage” can occur where seemingly cleavable sites are skipped. Extensive analysis of these “missed cleavages” has been studied by Keil [141], for which the most common are summarised in Table 2-1.

These tryptic missed cleavages, as well as the frequency of Arg/Lys residues in a protein, can sometimes make this mode of digestion unsuitable for producing peptides that are suitable for MS, or give poor protein sequence coverage. In these cases, digestion with alternative endoproteases should be investigated. Table 2-2 summarises some common endoproteases and their cleavage conditions. Once peptides with unique aa sequences are selected, the suitability of MS characteristics (e.g. sensitivity and stability) are considered, so as to ensure the best selection of surrogate peptides for LC-MS methods.

Table 2-1: Likely trypsin missed cleavage sites as defined by Keil [141]. Where P1 represents the primary binding site (specific for Arg (R) or Lys(K)), P1’-P3’ and P2-P4 represent the carboxyl-and amino-terminal secondary binding site carboxyl-terminal to P1, respectively.

P4	P3	P2	P1	P1’	P2’	P3’
			[RK]	[P]		
			[RK]	[RK]		
		[DE]	[RK]			
			[RK]	[DE]		
			[RK]		[DE]	[DE]
[DE]	[DE]		[RK]			
	[DE]		[RK]		[DE]	

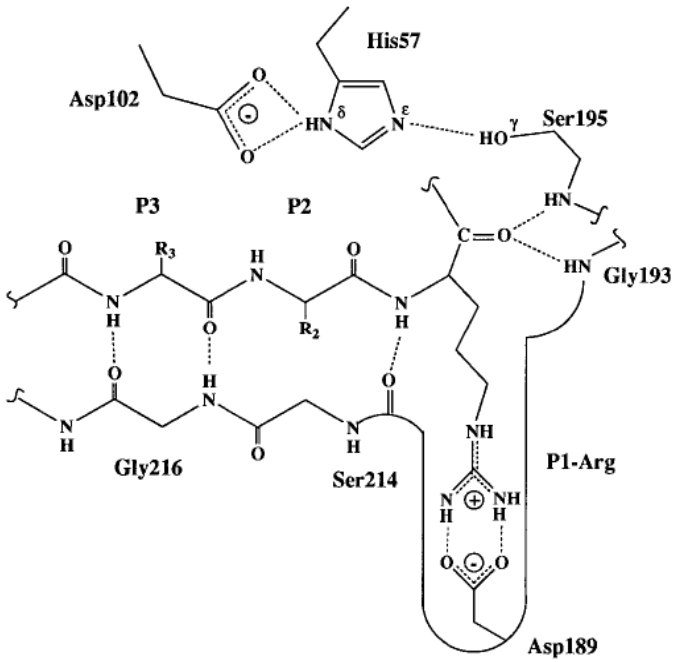


Figure 2.2-1: Showing the specificity of the Asp-189 residue at the S1 (P1) binding pocket of trypsin for Arg and Lys. Through nucleophillic attack from trypsin's Ser-195 to the carboxyl-terminal amide bond of the bound Arg/Lys, a catalytic cascade involving His-57 and Asp-102 and water forms an acyl-enzyme intermediate to result in the hydrolysis of the peptide bond. (Image taken with permission from [142]).

Table 2-2: Common digestion endoproteases used in proteomics and their cleavage specificities

Endoprotease	Cleavage Specificity	Optimal pH	Examples of use
Trypsin	carboxyl-terminal of R and K	7.8-8.7	Most commonly used
Lys-C	carboxyl-terminal of K	7.0-9.0	Produce larger peptides than trypsin; however sometimes used with trypsin to reduce K "missed cleavage"
Glu-C	carboxyl-terminal of E and to a lesser degree D	4.0-9.0	Used as an alternative to trypsin to produce different peptides and expand protein coverage. Cleavage sites are buffer specific.
Asp-N	amino-terminal of D	4.0-9.0	Sometimes used to complement Glu-C digests to produce similar peptides (1 aa longer)
Chymotrypsin	carboxyl-terminal of Y, F, W and L	7.0-9.0	Digestion of hydrophobic proteins (e.g. membrane proteins). Often produces small peptides.

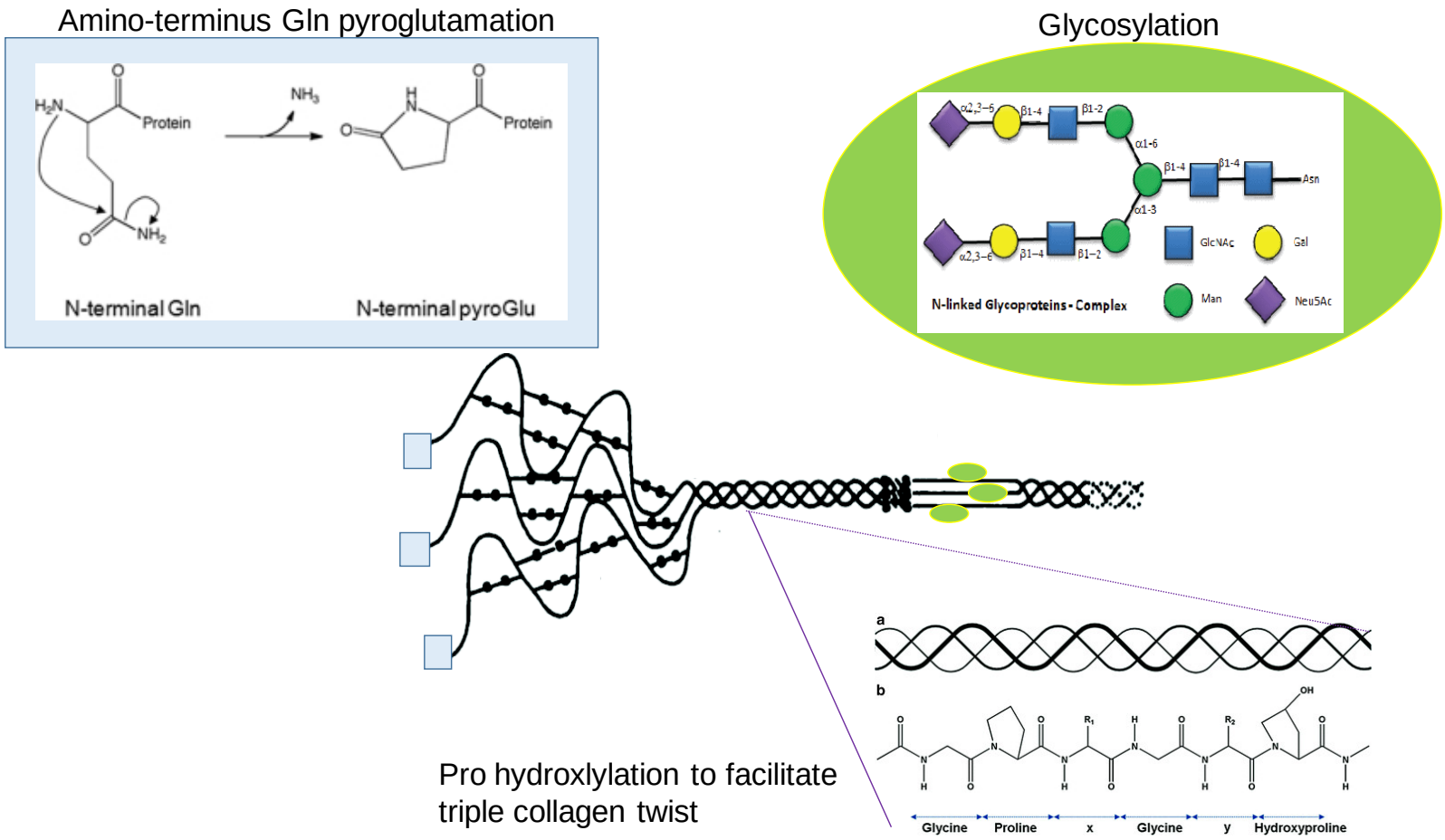


Figure 2.2-2: Illustration of P-III-NP highlighting PTMs determined from the topology and aa sequence. Where the shapes represent the location of modification in the protein structure.

2.3 **Materials and Methods**

All water used was purified using an Elgastat Pure Lab Option water purification unit and had ~18 MΩ cm resistivity. Analar NormaPur. Formic acid (99-100 %) was purchased from VWR Chemicals (Fontenay-sous-Bois, France). Ammonium bicarbonate (≥99 %, Reagent Plus®), Substance P acetate salt hydrate (≥95%, HPLC grade), and albumin solution human (HSA, 30 % in 0.85 % sodium chloride, protease free), were purchased from Sigma Aldrich (St. Louis, MO, USA).

2.3.1 ***In silico digestion of human P-III-NP***

Theoretical (*in silico*) digestion of human P-III-NP (UniProtKB entry E7ENY8) with trypsin, LysC, Chymotrypsin, Asp-N, and Glu-C was performed using ExPasy Peptide Cutter [143] and USCF Protein prospector [144], where zero missed cleavage was allowed for each digestion. To assess the uniqueness of *in silico* derived peptides (> 500 Da and > 5 aa), sequences were searched against the UniProtKB database using the Basic Local Alignment Search Tool (BLAST) with an expectation value (E) threshold of 10.

2.3.2 ***P-III-NP in silico derived peptides***

In the absence of international reference material for human P-III-NP, five *in silico* proteolytic derived P-III-NP peptides (Table 2-3 below) were synthesized by Fisher Scientific (Offenbach, Germany) for the purpose of developing LC-MS methods to target P-III-NP.

At the time of synthesis, it was unclear as to whether the highlighted amino acids in Tyr_6 (NGDPGIPGQPGSPGSPGPPGICESCPTGP***QNYSP)*** were included in circulating P-III-NP (personal communication (Prof. Cowan (supervisor) and Dr. L Bowers 2014 (Partnership for clean competition) and so were omitted from the synthesized peptide.

2.3.2.1 Sample Preparation of peptides for LC-MS

Sample preparation for LC-MS analysis was dictated by the presence of Cys (C) residues in the peptide amino acid sequence.

Table 2-3: Synthesized P-III-NP (human) *in silico* derived peptides

Endoprotease	Peptide reference	Peptide sequence
Trypsin	Tyr_1	QQEAVEGGCSHLGQSYADR
	Tyr_5	GDPGPPGIPGR
	Tyr_6	NGDPGIPGQPGSPGSPGPPGICESCPTGP (*QNYSP*)
GluC (*S. aureus* V8, phosphate)	GluC_2	GGCSHLGQSYAD
	GluC_10	CCAVCPQPPTAPTRPPNGQGPQGPKGD

2.3.2.1.1 Preparation of peptide standards

For Tyr_5, the only peptide not containing a Cys residue, pure peptide standards (1, 7, 14, 70, 140 and 210 nM) were prepared in formic acid (0.3 %) and acetonitrile (5 %) solution. By LC-MS/HRAMS, standards were used to develop a product scan method for Tyr_5, and establish the limit of detection (LOD) for the peptide as detailed in section 2.3.3.

2.3.2.1.2 Preparation of peptide standards (C derivatisation)

For all other peptides (containing Cys residues within the peptide sequence) spontaneous coupling of thiol residues to form inter- and intra-covalently linked disulphide bonds was expected. These bonds complicate LC-MS analysis, requiring the use of deconvolution software to interpret data. To avoid this step, disulphide bonds were reduced from our peptide standards using dithitohreitol (DTT. Bio-Rad Laboratories, Herts, UK) and later blocking by alkylation with iodoacetamide (IDA. Sigma Aldrich, St. Louis, MO, USA) to carbamidomethylate the C residue. Pure standards (1, 7, 14, 70, 140 and 210 nM) for each peptide were prepared in ammonium bicarbonate (100 mM) and derivatised using the following protocol:

1. In a 1.5 mL Lobind Eppendorf tube (Sigma Aldrich, Hamburg, Germany), DTT (100 mM in 100 mM NH_4HCO_3, 5 µL) was added to the sample (100 µL)and

vortexed (2-3 sec), prior to mixing at 900 rpm in a Thermomixer® for 10 min at 60 ˚C.

2. Samples were left to stand for a further 50 min at 60˚C in the Thermomixer®, before cooling to room temperature and centrifuging at 500 rpm for 5 min to collect condensate.
3. IDA (250 mM in 100 mM ammonium bicarbonate, 5 µL) was then added to the sample which was incubated in the dark for 30 min at room temperature.

Once derivatised, standards were prepared for LC-MS/HRAMS analysis by adding formic acid (2 %, 10 µL) to the sample. As with Tyr_5, standards were again used to develop product ion scans and establish LODs.

2.3.3 *Liquid chromatography coupled mass spectrometry-high resolution mass spectrometry (LC-MS/HRAMS)*

2.3.3.1 Liquid chromatography

A simple gradient LC method using a Thermo Fisher™ Dionex™ UltiMate™ 3000 UHPLC (ultra high performance liquid chromatography) system was developed to allow for sufficient separation of the *in silico* derived peptides (from a postulated P-III-NP digest) from potential matrix peptides. Using the settings detailed in Table 2-4 below, separation of sample (10 µL) was achieved using a Waters Acquity® UPLC® BEH 130Å column (C18, 1.7 µm, 2.1 x 50 mm), with mobile phases A (0.3 % formic acid in water) and B (0.3 % formic acid in acetonitrile) over 10 min.

Table 2-4: LC gradient conditions used to achieve separation of P-III-NP peptides by UHPLC

Time(min)	Flow(mL/min)	% A	% B
0.00	0.3	95	5
3.00	0.3	95	5
8.00	0.3	10	90
10.00	0.3	95	5

2.3.4 *Mass spectrometry*

2.3.4.1 MS tune conditions

To optimise MS source conditions suitable for analysing our peptides, substance P (1347 Da), a readily available but similar sized peptide was used. By infusing substance P (500 ng/mL, 371 nM) at 5 µL/min on a Thermo Fisher Scientific Q-Exactive with HESI II ionisation source in positive mode (using N_2 as source and collision gas), the following instrument parameters were established:

- sheath, auxiliary and sweep gas: 70, 13 and 0 arbitrary units
- capillary and heater temperature: 300 ˚C and 325 ˚C
- spray voltage: 3.75 kV

2.3.4.2 MS/HRAMS product scan methods

To select an appropriate precursor ion for each of the *in silico* derived P-III-NP peptides, a full scan spectrum was collected using high resolution accurate mass spectrometry (HRAMS) over the *m/z* scan range 200 - 3000. In HRAMS mode the resolving power was set to 140,000 FWHM, with an AGC target of 1×10^6 and a maximum injection time of 100 ms.

Using the full scan spectrum, singly-, doubly- and triply-charged ions (*m/z*) (refer to Table 2-5 below) for each peptide were extracted to identify the most abundant precursor. The selected precursor (± 1 amu) was used to trigger product ion scans in MS/HRAMS mode using normalised collision energies (nCE) 15, 20, 25 and 30 (%, arbitrary units) to select the best spectrum. Where complete or insufficient fragmentation of the precursor was observed (within the selected range), the nCE was appropriately adjusted.

Table 2-5: Peptide *m/z* values (singly-, doubly- and triply-charged ions) for *in silico* derived P-III-NP (human) peptides; where C represents carbamidomethylated (C_2H_3NO, +57.0215 Da) Cys residues.

Fragment	Peptide Sequence	$[M+H]^+$ (*m/z*)	$[M+2H]^{2+}$ (*m/z*)	$[M+3H]^{3+}$ (*m/z*)
Tyr_1	QQEAVEGGCSHLGQSYADR	2091.9149	1046.4611	697.9765
Glu-C_2	GGCSHLGQSYAD	1251.5059	626.2566	417.8401
Glu-C_10	CCAVCPQPPTAPTRPPNGQGPQGPKGD	2844.2974	1422.6524	948.7707
Tyr_5	GDPGPPGIPGR	1019.5268	510.2670	340.5138
Tyr_6	NGDPGIPGQPGSPGSPGPPGICESCPTGP	2743.2087	1372.1080	915.0744

2.3.4.3 Data-dependent mass spectrometry

To generate untargeted MS/HRAMS a product ion spectrum for the P-III-NP trypsin digest, the resolving power for the initial full (survey) scan by HRAMS was reduced to a resolution of 70,000 FWHM, to allow for the collection of more data points. Using a top 50 method and a precursor isolation window of ±1 amu, data dependent MS/HRAMS spectrum were collected across *m/z* 200-3000, with a resolving power of 17,500 FWHM, nCE 30, AGC target set to 5×10^4 and maximum injection time of 50 ms.

For data analysis, collected spectra were checked using the Matrix Science Mascot server (version 2.2.06) searching against all taxonomic entries in UniProtKB_sprot_130200. A peptide and fragment tolerance of 5 ppm and ±0.50 Da with a maximum of three missed cleavages was allowed, with a peptide charge 1+, 2+ and 3+ and above. Carbamidomethyl (C) was entered as a fixed modification, whilst Q-> pyroE (N-term Q), Oxidation (P) and Deamidated (NQ) were entered as variable modifications. Data was also checked with Carbamidomethyl (C) as a variable modification. Samples were analysed with and without the error tolerance tool.

2.3.5 *In vivo trypsin digestion of bovine P-III-NP*

To evaluate the yield of the synthesized *in silico* derived tryptic peptides from P-III-NP, an *in vivo* digestion of bovine P-III-NP (97 % homology with human, Institute of Bioanalytics, L.L.C (Branford, USA)) was evaluated by LC-MS. Using our developed methods full scan (HRAMS), product scan (MS/HRAMS targeting trypsin peptides) and data dependent MS/HRAMS spectrum were collected. Trypsin digestion of a P-III-NP

positive control (~100 µg/mL in 0.1 % human serum albumin (HSA)) and HSA negative control (0.1%) was performed as follows:

1. Using the protocol described in Section 2.3.2.1 above, each control (100 µL) was derivatised, to remove disulphide bonds.
2. Trypsin (100 µg, sequencing grade modified (Promega, Wisconsin USA)) was reconstituted by adding acetonitrile (50 mM, 20 µL) and then ammonium bicarbonate (100 mM, 480 µL) before repeatedly pipetting up and down (~6 times), to make a stock solution.
3. To each control, trypsin stock solution (2 µg, 10 µL) was added and vortexed before incubating overnight (min 8 h) at 37 ˚C whilst mixing at 1,400 rpm in a Thermomixer®.
4. To end trypsin digestion, formic acid (2 %, 10 µL) was added to each control and vortexed (2-3 sec) before evaporating at 45 ˚C in a Speedvac® (~3.5 h).
5. The controls were reconstituted in formic acid (0.3 %) and acetonitrile 5 % solution (50 µL) and transferred to an autosampler vial to be analysed by LC-MS.

2.4 **Results and Discussion**

2.4.1 ***In silico digestion of human P-III-NP***

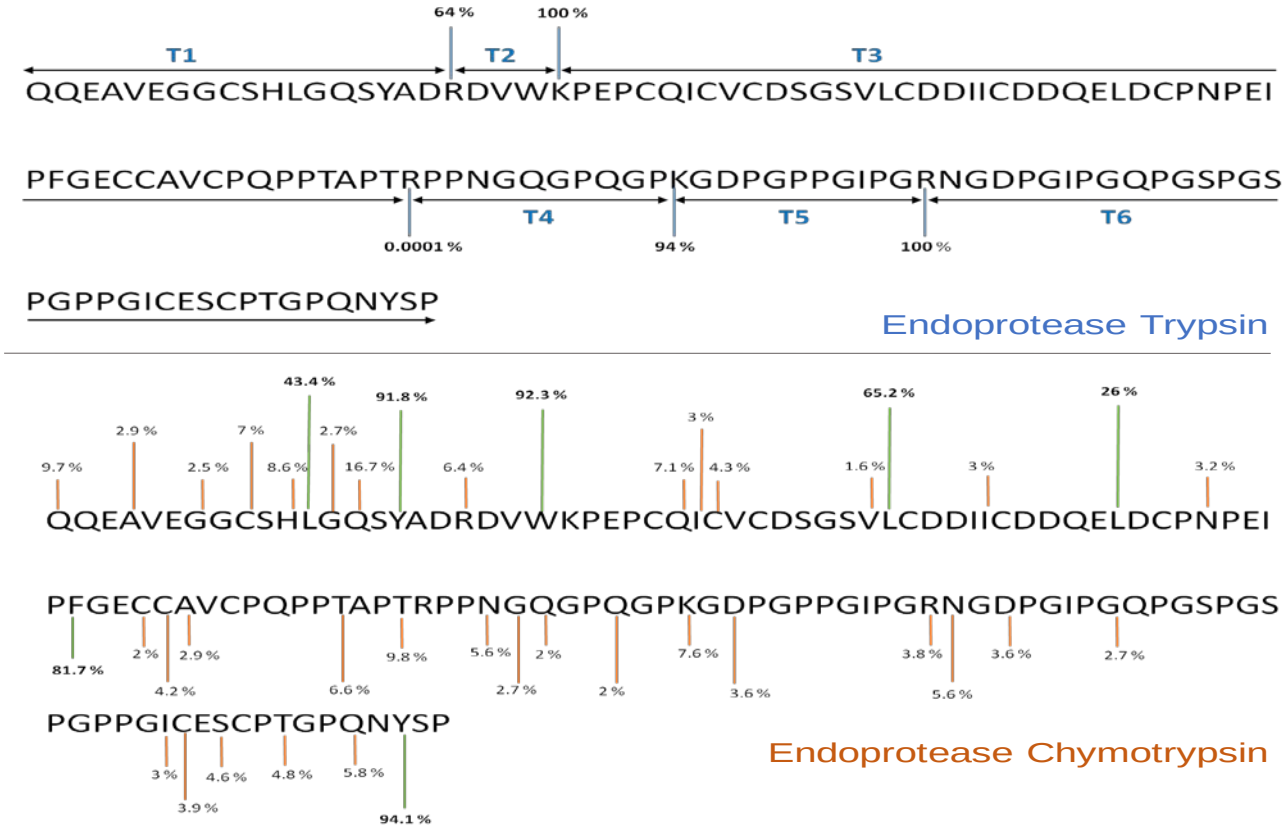


Figure 2.4-1: Map of cleavage sites after P-III-NP (human UniProtKB entry E7ENY8, monomeric) tryptic and chymotryptic digestion. The cleavage specificity (% cleavage) is a theoretical value based on a statistical treatment of the accumulated information for both endoproteases (for more details please see [141]).

Table 2-6: Showing the *in silico* digestion products for P-III-NP (human) for five endoproteases (trypsin, Lys-C, Chymotrypsin, Asp-N and Glu-C) with no missed cleavage allowed. Only fragments with > 5 aa and > 500 Da are shown. Peptide sequences highlighted in bold are unique to P-III-NP (all species) according to BLAST search against the UniProtKB database.

Endoprotease	Position of cleavage site	Peptide Reference	Peptide sequence (> 5 aa & 500 Da)	Peptide length (aa)	Molar Mass (Da)
Trypsin	1-19	Tyr_1	**QQEAVEGGCSHLGQSYADR**	19	2034.893
	20-85	Tyr_2-4	**DVWKPEPCQICVCDSGSVLCDDIICDDQELDCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPK**	66	7021.082
	86-96	Tyr_5	**GDPGPPGIPGR**	11	1019.527
	97-130	Tyr_6	**NGDPGIPGQPGSPGSPGPPGICESCPTGPQNYSP**	34	3218.415
Lys-C	1-23	Lys-C_1	**QQEAVEGGCSHLGQSYADRDVWK**	23	2563.163
	24-85	Lys-C_2	**PEPCQICVCDSGSVLCDDIICDDQELDCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPK**	62	6492.812
	86-130	Lys-C_3	**GDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCPTGPQNYSP**	45	4218.924
Chymotrypsin	1-12	Chym_1	**QQEAVEGGCSHL**	12	1257.553
	17-22	Chym_3	ADRDVW	6	761.358
	23-38	Chym_4	KPEPCQICVCDSGSVL	16	1677.764
	39-49	Chym_5	CDDIICDDQEL	11	1281.497
	50-58	Chym_6	DCPNPEIPF	9	1031.450
	59-128	Chym_7	**GECCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICESCPTGPQNY**	70	6703.070
Asp-N	1-17	Asp-N_1	**QQEAVEGGCSHLGQSYA**	17	1763.765
	20-32	Asp-N_3	DVWKPEPCQICVC	13	1519.674
	33-39	Asp-N_4	DSGVLC	6	593.260
	50-86	Asp-N_9	**DCPNPEIPFGECCAVCPQPPTAPTRPPNGQGPQGPKG**	37	3756.703
	87-98	Asp-N_10	**DPGPPGIPGRNG**	12	1133.570
	99-130	Asp-N_11	**DPGIPGQPGSPGSPGPPGICESCPTGPQNYSP**	32	3047.351
GluC (*S. aureus*V8, phosphate)	7-18	Glu-C_2	**GGCSHLGQSYAD**	12	1194.484
	21-33	Glu-C_4+5	VWKPEPCQICVCD	13	1519.674
	34-40	Glu-C_6	SGSVLCD	7	680.292
	61-119	Glu-C_10-12	**CCAVCPQPPTAPTRPPNGQGPQGPKGDPGPPGIPGRNGDPGIPGQPGSPGSPGPPGICE**	59	5569.626

In general, peptides < 500 Da and 5 aa in length were not unique to P-III-NP when searched with BLAST against UniProtKB. Hence, these peptides were excluded from the peptide list presented in Table 2-6 above. Of the endoproteases used to digest P-III-NP all the peptides produced by trypsin and LysC were found to be unique.

2.4.1.1 Trypsin

For trypsin, basic cleavage principles dictate hydrolysis of the amide bond carboxyl-terminal to Arg (R)/Lys (K) -X (where X ≠ Pro) to give four digest peptides for P-III-NP (Table 2-6). Being the most commonly used endoprotease to digest proteins, trypsin activity for all applicable amino acid tri-peptidic combinations has been documented by Keil [141], allowing the prediction of cleavage probability for P-III-NP as shown in Figure 2-3.

Considering trypsin's preference for Arg residues (compared to K), the 64 % probability predicted at the first cleave site, ^{19}R, is considerably low. This is attributed to secondary specificity of neighbouring negatively charged Asp (D) residues in positions P2 and P1' at the cleavage site, which can form salt bridges with Arg, thus, competes with ^{189}D in the trypsin S1 pocket and potentially inhibits enzyme activity. Although the probability of cleavage is lower than expected, extensive exposure to the enzyme should increase the yield of Tyr_1 from P-III-NP, hence we considered this peptide for developing LC-MS methods.

Amongst the *in silico* derived peptides, Tyr_2-4 (66 aa with mass 7021 Da) is the heaviest; this is an atypical product for trypsin which normally produces peptides of 7-35 aa (600-4000 Da) [140]. The size of the peptides is attributed to the presence of two "missed cleavage" sites at ^{21}V-W-**K**-P-E and ^{81}P-T-**R**-P-P within the aa sequence; although "zero missed cleavage" were permitted for *in silico* digest.

It is interesting to note that for the Tyr_2-4 "missed cleavage" site, ^{21}V-W-**K**-P-E, advanced cleavage specific data for trypsin [141] predict hydrolysis of this bond with a probability of 100 % (see Figure 2-3). This is due to trypsin's strong preference for cleavage at W-K-X bonds, including X=Pro. The initial "missed cleavage" predicted at

this site is dependent on the Pro residue in position P1', which is expected to hinder attachment of the protein to the specificity pocket, however, in this case the effect of Pro is secondary to that of Typ (W). Trypsin will therefore cleave Tyr_2-4 to two peptides ^{20}DVW**K** (Tyr_2) and ^{24}PEPCQICVCDSGSVLCDDIICDDQELDCPNPEIPFGE CCAVCPQPPTAPT**R**PPNGQGPQGPK (Tyr_3-4). Unfortunately Tyr_2 is not unique to P-III-NP and does not improve peptide yield.

For the second "missed cleavage" site, ^{81}P-T-**R**-P-P, using the advanced data set the probability of cleavage is 0.0001%. The Arg residue is neighboured by Pro residues in positions P3, P1' and P2'. The combined conformational rigidity of Pro at these positions will twist the peptide backbone and restrict the protein entering the trypsin specificity pocket to result in a "no cleavage" site. The remaining peptide, Tyr_3-4 (62 aa with mass 6492 Da), is still large for analysis using current MS instruments (maximum *m/z* 2000), which to observe would require a charge state ≥ 4, hence it is not considered as a surrogate peptide for method development.

The probability of cleavage for the other peptides, Tyr_5 and Tyr_6 is > 90 %, which is sufficient for good repeatability of peptide yield. Both are unique to P-III-NP, and although the Tyr_6 is larger than that expected for trypsin peptide (3218 Da), both are considered for developing LC-MS methods.

2.4.1.2 Lys-C

Lys-C cleaves carboxyl-terminal Lys(K)-X (X= any aa) bonds to give three unique peptides from P-III-NP. Unlike trypsin, Lys-C has high specificity for Lys-Pro, Lys-Arg or Lys surrounded by Asp/Glu in the P4-P2 and P1'-P3' of the cleavage site. Hence, to decrease Lys-affiliated trypsin "missed cleavages", a Lys-C digest can be included to increase peptide yield [145]. Besides ^{22}W-**K**-P, which already has a high affinity for hydrolysis by trypsin, none of these bonds exist in P-III-NP.

Lys-C generated peptides are potentially more basic than those produced by trypsin as they may contain R residues in addition to terminal Lys. This can facilitate higher charge states that can be beneficial for MS analysis. This is the case for Lys-C_1, a

fairly sized peptide (23 aa with mass 2563 Da) only requiring a charge state ≥ 2 to be visualised by conventional MS, which is easily achieved by one basic residue.

Lys-C_2 and Lys-C_3 are large peptides (>4000 Da) identical to Tyr_3+4 and combined Tyr_5 and Tyr_6, respectively. Lys-C_2 was earlier disregarded as a surrogate for P-III-NP and will be discussed no further. Lys-C_3 requires a charge state ≥ 3 for MS analysis which may be difficult to achieve with one basic residue at ^{96}R, hence it is also not considered for developing LC-MS methods.

2.4.1.3 Chymotrypsin

Chymotrypsin cleaves carboxyl-terminal Trp (W), Tyr (Y) and Phe (F) with high specificity and to a lesser extent at Leu (L), Met (M) and His (H) (in preferential order) [141] to give two unique P-III-NP peptides. Due to the frequency of cleavage sites, chymotrypsin derived peptides are typically small (i.e. < 5 aa). Hence it is interesting to observe Chym_7 (70 aa with a mass 6703 Da) amongst the products for this endoprotease. The size of this peptide may be attributed to the predominant tri-peptidic Gly-Pro-Pro repeat within the Col_3 domain of P-III-NP, from which the majority of the peptide originates.

The frequency of chymotrypsin cleavage for polypeptidic substrates has been studied in great detail [141] indicating additional low specificity chymotrypsin cleavage sites in P-III-NP at Gly (Q), Ala (A), Gly (G), Cys (C), Arg (R), Ile (I), Val (V) and Asn (N) residues as shown in Figure 2-3. For Chym_1, occurrence at these low specific sites could produce peptides: ^{1}Q, ^{2}QEA, ^{5}VEG, ^{8}GC, ^{10}SH and ^{12}L, which would no longer be unique to P-III-NP. Likewise, additional cleavage of Chym_7 can occur to produce two unique peptides ^{87}PGPPGIPGR and ^{105}QPGSPGSPGPPG.

The low selectivity of chymotrypsin for P-III-NP, could lead to a yield of 41 peptides. However, because of the typical low cleavage probability for most sites (only three sites > 90 %) (Figure 2-3) numerous missed cleavage are expected, which will affect the reproducibility of peptide yield from a digest. It is for this reason that chymotryptic peptides were not considered further for method development.

2.4.1.4 Asp-N

Asp-N cleaves amino-terminal Asp (D) to produce four unique peptides from P-III-NP. At high concentration additional amino-terminal cleavage of Glu (E) can be expected [146]. Occurrence of cleavage at these low specific sites affects the stability of Asp-N derived peptides, to potentially observe these additional peptides from:

Asp-N_1: ^{1}QQ and ^{3}EAVEGGCSHLGQSYA,
Asp-N_9: ^{50}DCPNP, ^{55}EIPFG and ^{60}ECCAVCPQPPTAPTRPPNGQGPQGPKG
Asp-N_11: DPGIPGQPGSPGSPGPPGIC and ESCPTGPQNYSP.

These additional cleavage sites will affect the reproducibility of peptide products from Asp-N digests, hence they are not considered further for method development.

2.4.1.5 Glu-C

In phosphate buffer, Glu-C will cleave amino-terminal Glu (E) and Asp (D) [147] to produce two unique peptides from P-III-NP. Cleavage probability is decreased with Pro residue in positions P3, P1' or P2'. This is evident for Glu-C_10-12, which has two missed cleavage sites at ^{85}K-G-**D**-P-G. and ^{97}N-G-**D**-P-G. Unlike trypsin and chymotrypsin, the nature of this endoprotease is not as well defined, hence the likelihood of occurrence of these Pro defined "missed cleavages". If we assume that full cleavage of the protein will occur, from Glu-C_10-12 we obtain additional unique peptides:

Glu-C_10: CCAVCPQPPTAPTRPPNGQGPQGPKGD (2672.2257 Da)
Glu-C_11: PGPPGIPGRNGD (1132.5625 Da)
Glu-C_12: PGIPGQPGSPGSPGPPGICE (1799.8512 Da)

The reproducibility of these peptides will be decreased because of the effect of Pro, however, we investigated Glu-C_10 as a possible surrogate peptide for P-III-NP because it provides additional sequence information to previously selected peptides Tyr_1, Tyr_5 and Tyr_6. It is ideal to have at least two peptides from digestion for LC-MS analysis, hence, Glu-C_2 was also investigated.

2.4.2 *MS-HRAMS of in silico derived peptides*

2.4.2.1 Trypsin *in silico* derived peptides

The doubly- and triply-charged peptides ion for Tyr_1 were observed in full scan mode, with the triply charged ion being the most abundant. Using a normalised collision energy of 18 a product ion scan spectrum for the $[M+3H]^{3+}$ ion (*m/z* 697.98 ±0.2 amu) was obtained (Figure 2-4 below). The most abundant peptide fragment in the product ion scan was the deamidated derivative of the precursor ion, $[M-NH_3+3H]^{3+}$. Under low energy MS/MS conditions, deamidated peptides can result from charge directed ammonia loss from the side chain of Gln, Glu, Arg or Lys residues. For Tyr_1,the presence of b_z-NH_3 series (especially b_2-NH_3) suggests the site of ammonia loss is the amino-terminus Gln. It can be argued that in-solution spontaneous pyroglutamation of Gln can also produce these fragments, however, full scan Tyr_1 spectrum show the deamidated $[M+3H]^{3+}$ derivative is < 10 % of its unmodified precursor.

For Tyr_5 all the precursor ions were observed in full scan, with the doubly-charged ion being the most abundant. Using a normalised collision energy of 18, a product ion scan of the $[M+2H]^{2+}$ ion (510.57 ± 0.2 amu) was obtained (Figure 2-5 below). The most abundant peptide in the product scan was y_9^{2+}, resulting from the cleavage of the ^{2}D-^{3}P amide bond, which is unusually labile at acidic pH [148]. The higher proton affinity of Pro to Asp, allows the y- fragment to retain the "mobile" proton to give the observed double charge. In source production of this peptide was investigated in full scan mode, for which the y_9^{2+} peptide was observed < 10 % of the intact peptide.

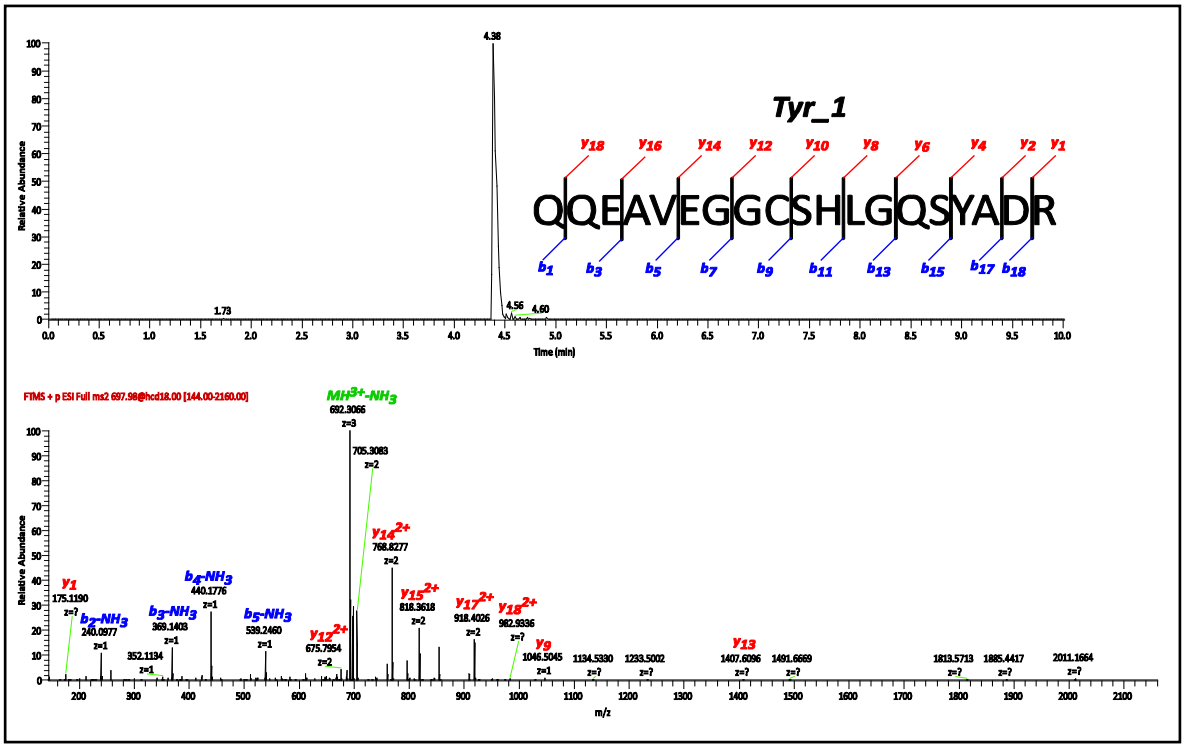


Figure 2.4-2: Extracted ion chromatogram at *m/z* 697.9765 ±5 ppm and corresponding product scan spectrum of the fragmented carbamidomethylated Tyr_1 triply charged peptide using a normalised collision energy of 18. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged.

A prominent y- series is observed for Tyr_5, resulting from proton sequestering at the carboxyl-terminus Arg. Cleavage of the Pro amide bonds appear to dictate the peptide fragments observed in the product ion scan spectrum (Figure 2-5) for Tyr_5, where y_9^{2+}, y_7^{+} and y_3^{+} ions are the most abundant ions. This is to be expected, as Pro residues have been shown to have a bias for amino-terminal fragmentation due to the structural hindrance imposed on the aa for nucleophillic attack of the amino-terminal carbonyl [149].

Where a doubly-charged y- ion is observed, both the sequestered and "mobile" protons are retained by the truncated y-peptide ion. For singly-charged y- ions, the "mobile" proton is likely to be retained by the leaving b- ion. Low abundance of b- ions in the spectrum indicate instability of the ions formed which are further internally fragmented to form smaller mass b- ions or a- ions (e.g. b_3^{+} to a_3^{+} ion transition).

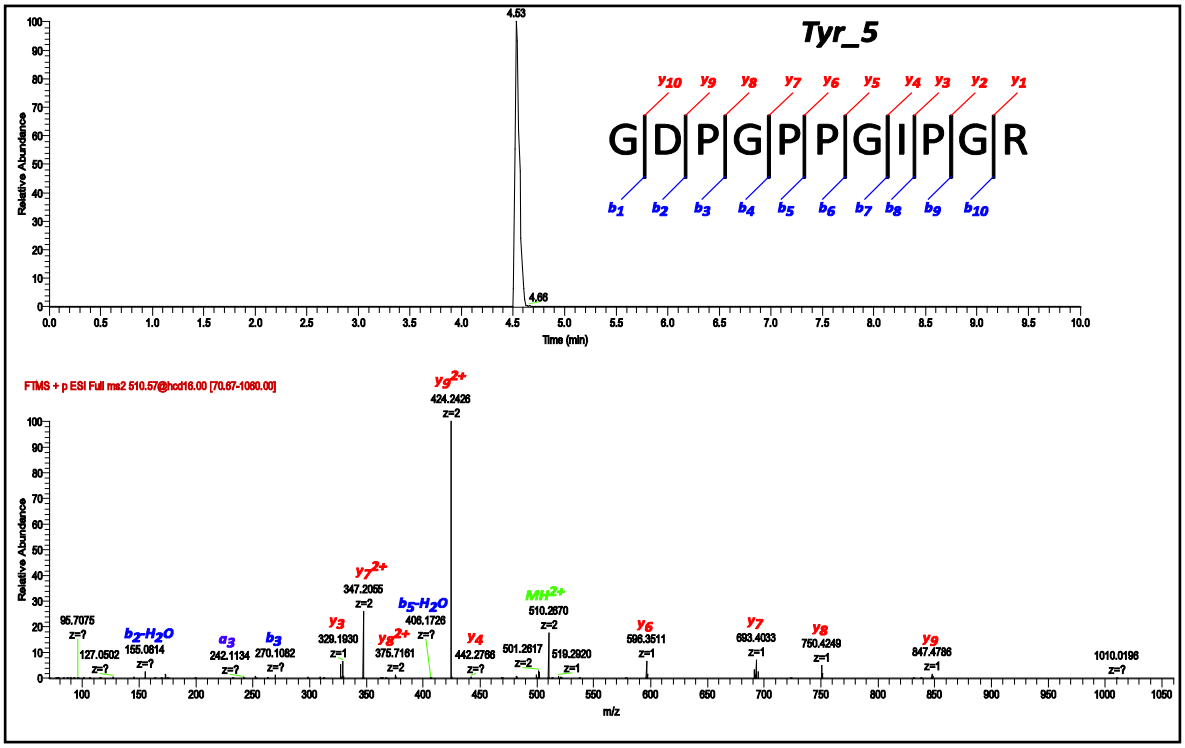


Figure 2.4-3: Extracted ion chromatogram at *m/z* 510.2671 ±5 ppm and corresponding product scan spectrum of the fragmented Tyr_5 doubly-charged peptide ion using a normalised collision energy of 18. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged.

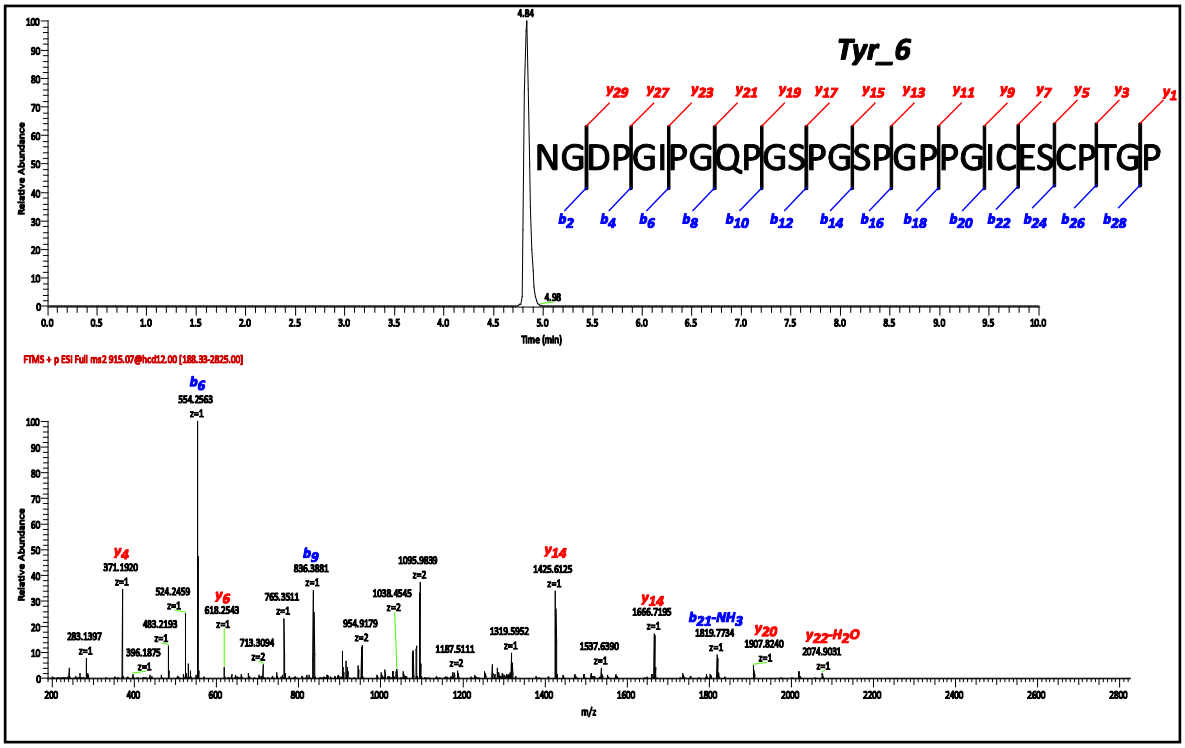


Figure 2.4-4: Extracted ion chromatogram at *m/z* 915.0744 ±5 ppm and corresponding product scan spectrum of the fragmented carbamidomethylated Tyr_6 triply charged peptide ionusing a normalised collision energy of 12. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged.

For Tyr_6, the doubly- and triply-charged peptide ions were observed in full scan, with the triply charged ion being the most dominant. Using a normalised collision energy of 12, a product scan spectrum of $[M+3H]^{3+}$ (*m/z* 915.07 ± 0.2 amu) was obtained (Figure 2-6 above), where the most abundant peptide fragment was b_6^+. Unlike the other two tryptic peptides, Tyr_6 does not have a terminal basic group, so it is not surprising that there is not a dominant y- series for this peptide. The b- ions seem to be more stable for this peptide, as they extend closer to the carboxyl-terminus thus providing greater peptide coverage (e.g. b_{21}-NH_3^+). Like Tyr_5 fragmentation of Tyr_6 is biased to the amino-terminus of Pro residues (e.g. b_6^+, b_9^+, y_{14}^+ and y_4^+). It is not surprising to observe mostly singly-charged peptides in the product scan spectrum for Tyr_6 as there are not highly basic aa in the peptide sequence that can sequester the available protons.

2.4.2.2 Glu-C *in silico* derived peptides

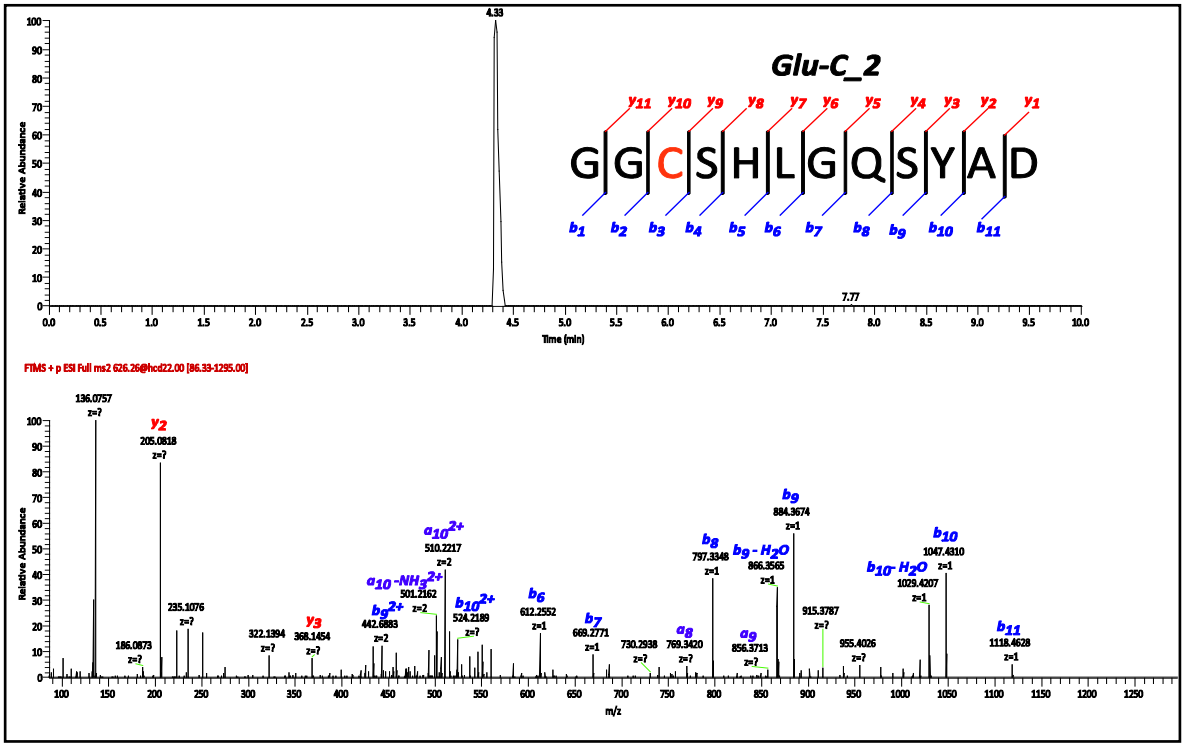


Figure 2.4-5: Extracted ion chromatogram at *m/z* 626.2566 ±5 ppm and corresponding product scan spectrum of the fragmented carbamidomethylated Glu-C_2 doubly-charged peptide ion using a normalised collision energy of 22. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged.

The doubly- and singly-charged peptide ions for Glu-C_2 were observed in full scan for which the doubly-charged ion was most dominant. Using a normalised collision energy of 22 a product scan spectrum of the $[M+2H]^{2+}$ (*m/z* 626.26 ± 0.2 amu) was obtained (Figure 2-7 above). Unlike the tryptic peptides, a predominant b- series is observed with full peptide coverage. Lack of a carboxyl-terminus basic group accounts for the low representation of y- ions in the spectrum. There is a basic His residue incorporated in most of the fragments, suggesting a possible site for proton sequestering, especially where doubly-charged fragment ions are observed (b_{10}^{2+} and b_9^{2+}). Instability of b_9 and b_{10} ions is evident as additional internal fragmentation and loss of neutrals is observed in the spectrum with the presence of $[b_9\text{-}H_20]^+$, a_9, $[b_{10}\text{-}H_20]^+$, a_{10}^{2+} and $[a_{10}\text{-}NH_3]^{2+}$.

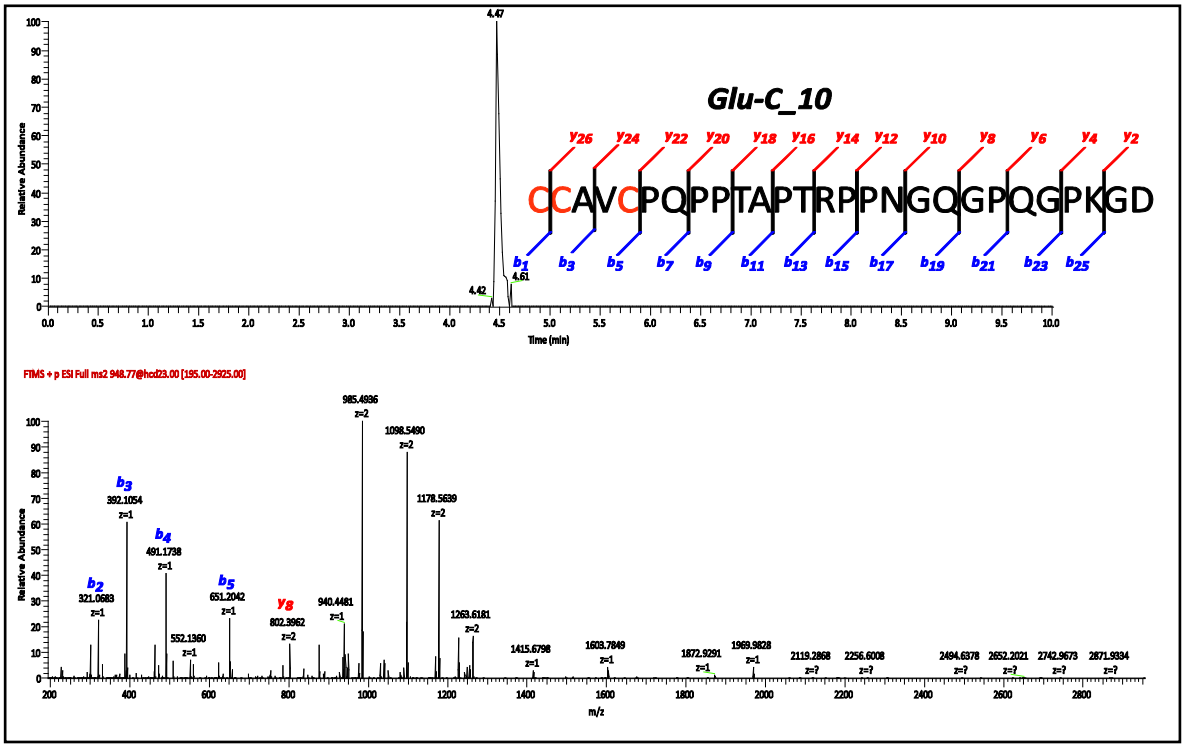


Figure 2.4-6: Extracted ion chromatogram at *m/z* 948.7707 ±5 ppm and corresponding product scan spectrum of the fragmented carbamidomethylated Glu-C_10 triply charged peptide ion using a normalised collision energy of 29. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged.

For Glu-C_10, the doubly- and triply-charged peptide ions were observed in full scan for which the triply charged ion was most dominant. Using a normalised collision energy of 29, a product ion scan of $[M+3H]^{3+}$ (*m/z* 948.7707 ±0.2 amu) was obtained (Figure 2-8 above). A poor quality spectrum was obtained for this peptide, which

required a relatively high energy (compared to the other peptides) for fragmentation. Glu-C_10 contains two basic residues within its aa peptide sequence, which can sequester available protons to yield multiple charged peptide fragments. At *m/z* 945.4936, 1098.5490 and 1178.5639 the more abundant isomers of y_{20}^{2+} (*m/z* 984.997), y_{22}^{2+} (*m/z* 1097.5536) and y_{23}^{2+} (m/ 1177.5689) are displayed.

2.4.2.3 Limit of detection

Table 2-7: Tentative limit of detection for *in silico* derived P-III-NP peptides (10 µL, determined by the three most dominant product ions being greater than 1 x 10^4) from scan data obtained by LC-MS/HRAMS.

Endoprotease	Peptide reference	Tentative Limit of Detection (LOD)/nM
Trypsin	Tyr_1	7
	Tyr_5	7
	Tyr_10	7
Glu-C	Glu-C_2	7
	Glu-C_10	14

As potential surrogate peptides for P-III-NP, determination of instrument sensitivity for each *in silico* derived peptide, by defining the limit of detection is necessary to determine the applicability for bioanalysis. Basal serum concentrations P-III-NP are 1-5ng/mL (23-117 pM, 100 µL) [52]. Using the developed product ion scan methods and the LOD determined (Table 2-7) a minimum P-III-NP concentration of 300 ng/mL (Tyr_1, Tyr_5, Tyr_6 and Glu-C_2) and 600 ng/mL (Glu-C_10) can be analysed. Combined trypsin produces more sensitive peptides for LC-MS, hence we perform *in vivo* digestion on P-III-NP to verify the production of *in silico* derived peptides from bovine foetal material. Sharing 97 % sequence homology to human P-III-NP, the following aa substitutions (highlighted in red) were made for bovine (UniProt_TrEMBL entry Q08E14) methods:

Tyr_1: QQEAVDGGCSHLGQSYADR
Tyr_6: NGDPGPPGSPGSPGSPGPPGICESCPTG*GQNYSP*

2.4.3 ***Partial characterisation of bovine P-III-NP by in vivo trypsin digestion***

Using the developed LC-MS product ion scans targeting Tyr_1 (bovine), Tyr_5 and Tyr_6 (bovine), the digest products of a trypsin bovine digest was analysed, for which

none of the peptides were found. Research into the possible P-III-NP post-translational aa modifications (PTM) that would alter the observed mass of the targeted peptides was undertaken.

2.4.3.1 Bovine Tyr_1 peptide

Deamidation of the exposed amino-terminus Gln to form an acidic *pyro*E (Q-*NH_3, Gln-NH_3*) derivative is possible; resulting in a -17.0265 Da shift in the mass of the target peptide. Being enzymatic and spontaneously driven, this modification is expected to be extensive or even complete for serum P-III-NP. Hence, it is not surprising to observe in the full scan MS data of the digest P-III-NP products doubly- and triply-charged ions corresponding to the *pyroE*_T1 (see Table 2-8 for reference values) peptide. Unlike Tyr_1, the doubly-charged peptide ion was the most abundant; this suggests that for Tyr_1 the amino-terminus Gln is an important protonation site.

Using a normalised collision energy of 30 a product ion scan spectrum of the carbamido-methylated $[M+2H]^{2+}$ PTM peptide (1030.94 ±0.2 amu) was obtained (see Figure 2-9 below). Evidence of Gln-*NH_3* is observed in the mass of the b- ions; which are identical to the b-NH_3 ions for the *in silico* derived Tyr_1 peptide (Figure 2-4 above). Full peptide coverage for the PyroE_T1 peptide was observed with overlapping y- (83 %) and b- (70 %) ions (Table 2-9). The y_7^+ (GQSYADR) peptide fragment at *m/z* 796.3566 (theoretical *m/z* 796.3584) is the most prominent in the spectrum. The largest peptide fragment to be observed is b_{18}^+ (Q_{-NH3}QEAVDGGCSHLGQSYA) at *m/z* 1886.7677 (theoretical *m/z* 1866.7610) and the smallest is b_2^+ (Q_{-NH3}Q) at *m/z* 240.0977 (theoretical *m/z* 240.0979).

Table 2-8: Bovine P-III-NP trypsin digestion products selected for MS analysis. The sequences shown are the modified *in silico* derived peptides and the expected products based on *a priori* knowledge of PTMs (highlighted in bold). For T1, deamidation of the amino-terminus Gln (Q) to Q_{-NH3} results loss of 17.0265 Da to the PTM peptide. For the T5 and T6 peptide, hydroxylation of Pro in the Gly-X-Pro results in the addition of 15.9949 Da for every P_{-OH} residue. For the T6, deamidation of the susceptible Amino-terminal Asn (N)-Gly to Asp (D)-G results in further addition of 0.984 Da to the PTM peptide. The theoretical singly-, doubly- and triply-charged ions (*m/z)* for each peptide is presented.

Peptide		Amino acid sequence	Peptide ion (*m/z* values)		
			$[M+H]^+$	$[M+2H]^{2+}$	$[M+3H]^{3+}$
Tyr_1	In silico	1QQEAVDGGCSHLGQSYADR19	2020.8777	1010.9425	674.2974
	Carbamidomethylated	1QQEAVDGGCSHLGQSYADR19	2077.8992	1039.4532	693.3046
	Carbamidomethylated PTM	1Q-NH3QEAVDGGCSHLGQSYADR19	2060.8727	1030.9400	687.6291
Tyr_5	In silico	86GDPGPPGIPGR96	1019.5269	510.2671	340.5138
	PTM	86GDP-OHGPP-OHGIP-OHGR96	1067.5116	534.2594	356.5087
Tyr_6	In silico	97NGDPGPPGSPGSPGSPGPPGICESCPTGGQNYSP130	3121.3262	1561.1667	1041.1136
	Carbamidomethylated	97NGDPGPPGSPGSPGSPGPPGICESCPTGGQNYSP130	3235.3691	1618.1882	1079.1279
	Carbamidomethylated PTM	97DGDP-OHGPP-OHGSP-OHGSP-OHGSP-OHGPP-OHGICESCPTGGQNYSP130	3332.3226	1666.6650	1111.4457

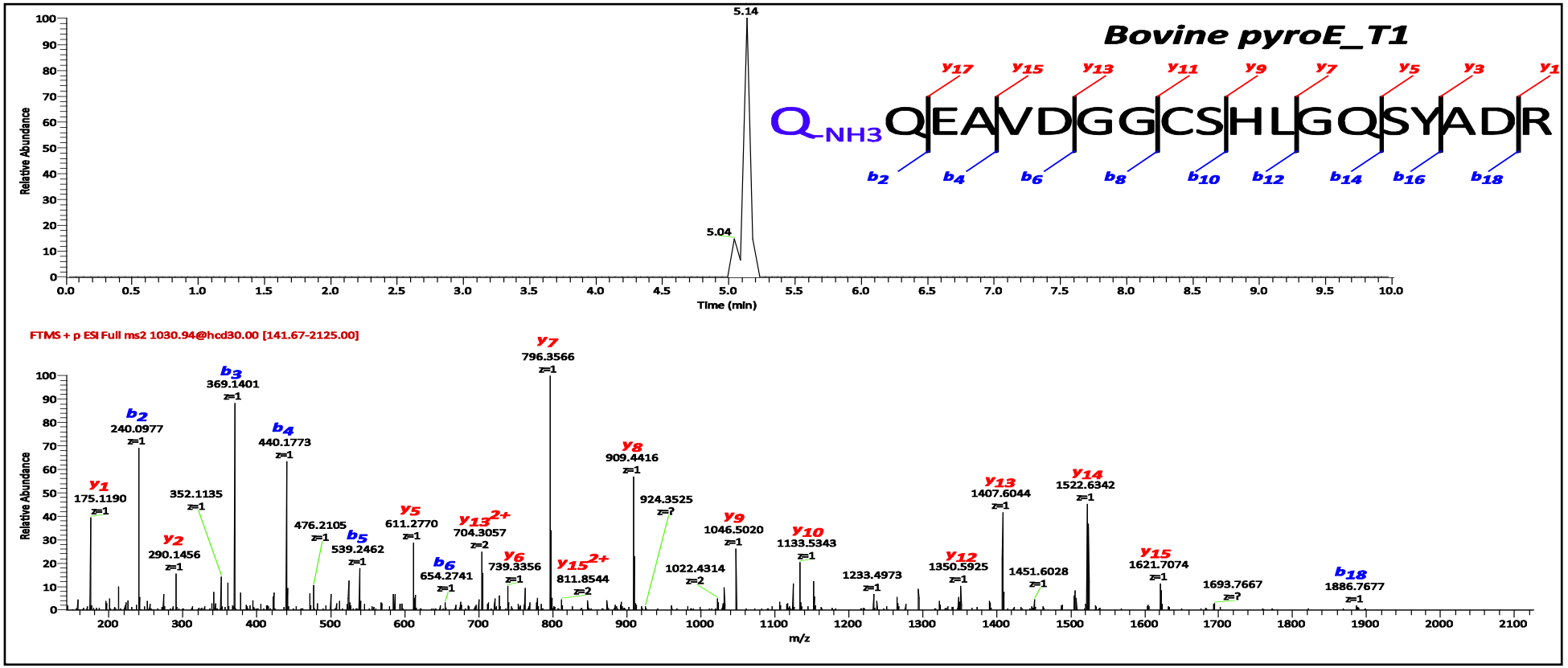


Figure 2.4-7: Extracted ion chromatogram at *m/z* 1030.9400 ±5 ppm and corresponding product ion spectrum of the fragmented post translational modified carbamidomethylated bovine *pyro*E_T1 doubly-charged peptide ion using a normalised collision energy of 30. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged (Refer to Table 2-9 below for theoretical *m/z* values).

Table 2-9: Possible b- and y-ions of *in vivo* tryptic bovine P-III-NP carbamidomethylated *pyro*E_T1 peptide, showing the fragment ions observed (in bold) from the extracted ion chromatogram shown in Figure 2-9 above and sequence coverage in green.

b- fragment				y- fragment		
m/z **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**	**fragment no.**	**amino acid sequence**	**fragment no.**	***m/z*** **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**
-	-	1	**Q-$_{NH3}$**	19	-	-
240.0979	-	2	**Q**	18	1949.8406	975.4240
369.1405	-	3	**E**	17	1821.7821	911.3947
440.1776	-	4	**A**	16	**1692.7395**	**846.8734**
539.2460	-	5	**V**	15	**1621.7023**	**811.3548**
654.2729	-	6	**D**	14	**1522.6339**	**761.8206**
711.2944	-	7	**G**	13	**1407.6070**	**704.3071**
768.3159	-	8	**G**	12	**1350.5855**	**675.7964**
928.3465	-	9	**C (+C_2H_3NO, +57Da)**	11	**1293.5641**	647.2857
1015.3786	-	10	**S**	10	**1133.5334**	**567.2703**
1152.4375	576.7224	11	**H**	9	**1046.5014**	**523.7543**
1265.5212	633.2644	12	**L**	8	**909.4425**	455.2249
1322.5430	661.7751	13	**G**	7	**796.3584**	398.6828
1450.6016	725.8044	14	**Q**	6	**739.3369**	370.1721
1537.6336	769.3204	15	**S**	5	**611.2784**	306.1428
1700.6969	850.8521	16	**Y**	4	**524.2463**	262.6268
1771.7340	886.3707	17	**A**	3	**361.1830**	181.0951
1886.7610	943.8841	18	**D**	2	**290.1459**	145.5766
-	-	19	**R**	1	**175.1190**	88.0631

2.4.3.2 Tyr_5 peptide

For Tyr_5 which originates from the collagen-like, Col 3, domain of P-III-NP, post-translational hydroxylation at the second P in the G-P-P repeat is expected; this will result in a +47.9847 Da shift in the mass of the target peptide. In full scan, the singly- and doubly-charged *hydroxy*P_T5 peptide ions were observed, for which the doubly-charged peptide was more abundant (see Table 2-8 for reference values). Using a normalised collision energy of 20, a product scan spectrum of the $[M+2H]^{2+}$ PTM peptide (534.26 ±0.2 amu) was obtained (see Figure 2-10 below).

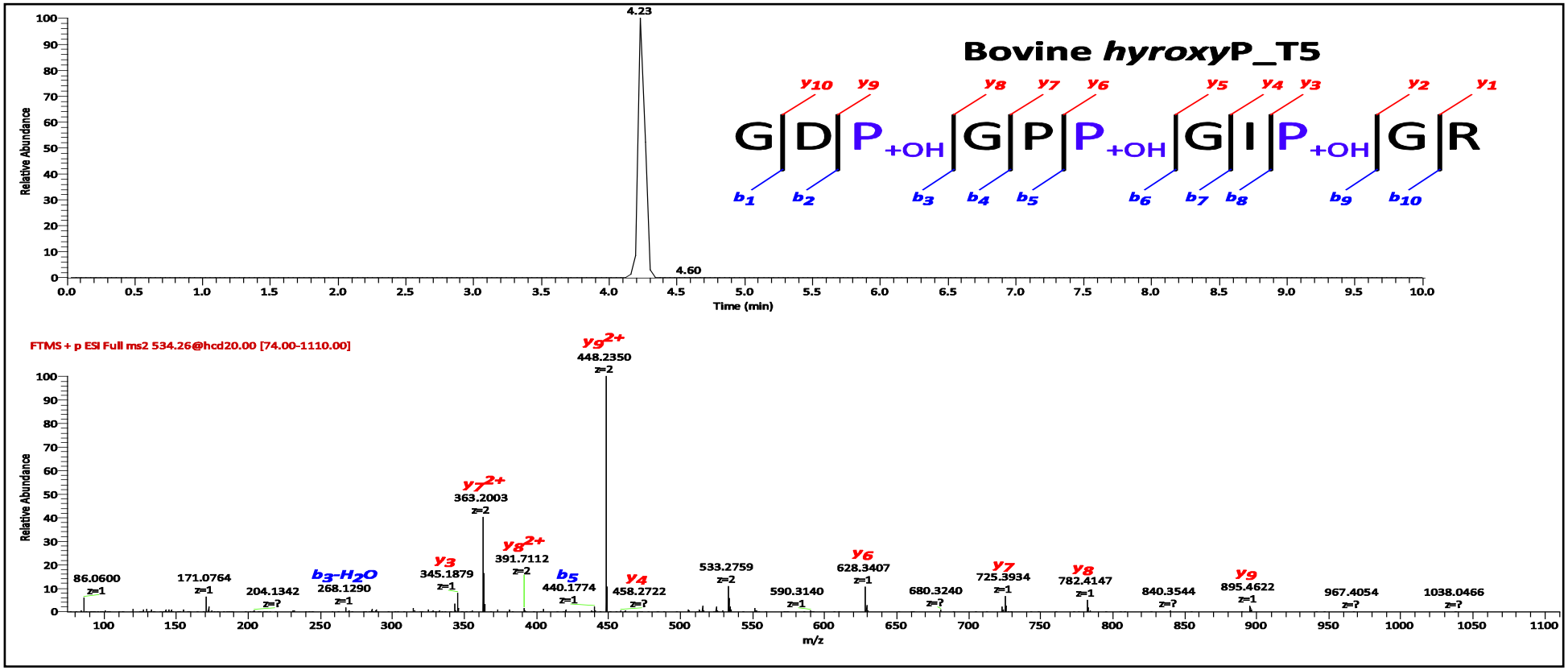


Figure 2.4-8: Extracted ion chromatogram at *m/z* 534.2594 ±5 ppm and corresponding product ion spectrum of the fragmented bovine P-III-NP post translational modified hydroxyP_T5 doubly-charged peptide ion using a normalised collision energy of 20. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged (Refer to Table 2-9 below for theoretical *m/z* values).

The observed mass of y_6^+ ($P_{+OH}GIP_{+OH}GR$) at *m/z* 628.3407 (theoretical *m/z* 628.3413) and y_7^+ ($PP_{+OH}GIP_{+OH}GR$) at *m/z* 725.3934 (theoretical *m/z* 725.3941) support the stereo and site specificity of hydroxylation for which we observe a mass difference of 97.0528 Da, which is equivalent to P. Full peptide coverage for *hydroxy*P_T5 was observed with 100 % y- ions and 80 % b- ions across the peptide sequence (Figure 2-10). The fragmentation pattern is similar to that of Tyr_5 (Figure 2-5) where y_9^{2+} ($P_{+OH}GPP_{+OH}GIHP_{+OH}GR$) at *m/z* 448.2350 (*m/z* 448.2352) is the most prominent in the spectrum, followed by y_7^{2+} (PP_{+OH} GIP_{+OH} GR) at *m/z* 363.2003 (theoretical *m/z* 363.2007). All other peptide fragment ions were much less abundant.

Table 2-10: Possible b- and y-ions of *in vivo* tryptic bovine P-III-NP hydroxyP_T5 peptide, showing the fragment ions observed (in bold) from the extracted ion chromatogram shown in Figure 2-10 above and sequence coverage in green.

b- fragment				**y- fragment**		
***m/z* $[M+H]^+$**	***m/z* $[M+2H]^{2+}$**	**Fragment no.no.**	**amino acid sequence**	**fragment no.**	***m/z* $[M+H]^+$**	***m/z* $[M+2H]^{2+}$**
-	-	1	**G**	11	-	-
173.0557	-	2	**D**	10	1010.4901	**505.7487**
286.1034	-	3	**P_{+OH}**	9	**895.4632**	**448.2352**
343.1248	-	4	**G**	8	**782.4155**	391.7114
440.1776	-	5	**P**	7	**725.3941**	**363.2007**
553.2253	-	6	**P_{+OH}**	6	**628.3413**	314.6743
610.2467	-	7	**G**	5	**515.2936**	258.1504
723.3308	-	8	**I**	4	**458.2722**	229.6397
836.3785	-	9	**P_{+OH}**	3	**345.1881**	173.0977
893.3999	-	10	**G**	2	**232.1404**	116.5738
-	-	11	**R**	1	**175.1190**	88.0631

2.4.3.3 Bovine Tyr_6 peptide

Among the trypsin peptides for P-III-NP; Tyr_6 is most susceptible to post translation modification. The first expected PTM is hydroxylation of Pro in the region of the peptide originating from Col 3, resulting in a +15.9949 Da mass shift for each modified P in the peptide. An additional -17.0265 Da mass shift for Tyr_6 can occur where spontaneous deamidation of the amino-terminus Asn to Asp is possible. With these modifications the aa sequence for PTM_T6 will be DGD***P-$_{OH}$***GP***P-$_{OH}$***GS***P-$_{OH}$***GS***P-$_{OH}$***GS***P-$_{OH}$***GP***P-$_{OH}$***GICESCPTGGQNYSP.

Potential *N*-linked and *O*-linked glycosylation at ^{126}**Q-N-Y**, ^{108}**S**, ^{111}**S**, ^{120}**S**, ^{123}**T** and ^{129}**S** have been identified within the peptide sequence for Tyr_6. However, lack of characterization of the oligosaccharide(s) and the number of occupied sites they occupy, makes it difficult to predict the adjusted peptide mass for Tyr_6. Cleaving these probable oligosaccharides from P-III-NP prior to trypsin digestion, would eliminate this unknown. However this step would complicate sample preparation procedures and increases peptide loss and is therefore not ideal.

A very small amount of non-glycosylated triply charged PTM_T6 peptide was observed at *m/z* 1111.4458 (theoretical *m/z* 1111.4457) from a digest of concentrated bovine P-III-NP (~ 100 µg/mL, 100 µL). Using a normalised collision energy of 20, a product scan of 1111.45 ± 2 amu was generated, for which 12 % b- ions and 9 % y- ions were observed (refer to Table 2-11 for details). Besides obtaining a poor quality spectrum for this peptide, the insufficient yield and unknown ratio of glycosylated and nonglycosylated peptide complicate quantification. This meant that PTM_T6 was no longer considered as a surrogate peptide for developing LC-MS methods.

Table 2-11: Possible b and y ions of *in vivo* tryptic bovine P-III-NP deamidated + hypdroxyP_T6 peptide, showing the fragment ions observed (in bold) from the extracted ion chromatogram, and sequence coverage in green.

b- fragment			**y- fragment**	
m/z **$[M+H]^+$**	**fragment no.**	**amino acid sequence**	**fragment no.**	***m/z*** **$[M+H]^+$**
-	1	**N_{-NH3} →D**	34	-
173.0557	2	**G**	33	3217.2957
288.0826	3	**D**	32	3160.2742
401.1303	4	**P+OH**	31	3045.2473
458.1518	5	**G**	30	2932.1996
555.2045	6	**P**	29	2875.1781
668.2522	7	**P+OH**	28	2778.1254
725.2737	8	**G**	27	2665.0777
812.3057	9	**S**	26	2608.0562
925.3534	10	**P+OH**	25	2521.0242
982.3748	11	**G**	24	2407.9765
1069.4069	12	**S**	23	2350.9551
1182.4546	13	**P+OH**	22	2263.9230
1239.4760	14	**G**	21	2150.8754
1326.5080	15	**S**	20	2093.8539
1439.5557	16	**P+OH**	19	2006.8219
1496.5772	17	**G**	18	1893.7742
1593.6300	18	P	17	1836.7527
1706.6776	19	P+OH	16	1739.7000
1763.6991	20	G	15	1626.6523
1876.7832	21	I	14	1569.6308
2036.8138	22	**C** **(+C_2H_3NO, +57 Da)**	13	**1456.5468**
2165.8564	23	**E**	12	1296.5161
2252.8884	24	**S**	11	1167.4735
2412.9191	25	**C** **(+C_2H_3NO, +57 Da))**	10	**1080.4415**
2509.9718	26	**P**	9	**920.4108**
2611.0195	27	**T**	8	823.3581
2668.0414	28	**G**	7	722.3104
2725.0624	29	**G**	6	665.2889
2853.1210	30	**Q**	5	608.2675
2967.1639	31	**N**	4	480.2089
3130.2273	32	**Y**	3	366.1660
3217.2593	33	**S**	2	203.1026
-	34	**P**	1	116.0706

2.4.3.4 Human serum albumin

As the most likely matrix to source P-III-NP for bio-analysis, possible interference from human serum albumin (HSA) for P-III-NP peptides was investigated. Analysis of a HSA control solution using PTM product ion scan methods for *pyroE*_T1 and *hydroxy*P_T5 was negative for the peptides. Hence we conclude that the tryptic peptides produced by HSA are different in mass from those generated from P-III-NP.

2.4.3.5 Data dependent analysis

Untargeted and *pyroE*_T1 targeted data-dependent spectrum for the bovine tryptic digest products returned negative results for P-III-NP when searched using Mascot against UniprotKB. It was later discovered that, unlike the human P-III-NP (UniProtKB entry E7ENY8), bovine P-III-NP(UniProt_TrEMBL entry Q08E14) is unreviewed and would not be included in the UniProt_SwissProt database that was applied to our data. As such, the data was reprocessed with the error tolerance tool enabled to allow for the expected aa substitutions between bovine and human variant; however, no significant score for P-III-NP was obtained.

2.5 **Summary**

Of the five endoproteases used to *in silico* digest P-III-NP (human, UniProtKB entry E7ENY8), trypsin produced more unique peptides suitable for mass spectrometry. In the absence of human P-III-NP international reference material, bovine P-III-NP (sharing 97 % sequence homology with man) was *in vivo* digested with trypsin prior to LC-MS analysis, to enable the verification of the *in silico* derived peptides. Initial analysis of digest products gave insufficient response for P-III-NP peptides, however with the consideration of potential PTM aa (and their effect on the expected peptide mass) modified variants of tryptic peptide, Tyr_1 *(pyroE*_T1) and Tyr_5 (*hydoxy*P_T5), were identified with 100 % sequence coverage. Hence, these peptides were targeted as surrogates of the P-III-NP peptides. They are the focus of Chapter 3, where they are used to assist in the development of highly sensitive LC-MS methods capable of identifying P-III-NP at basal serum levels.

Chapter 3 *: Analysis of P-III-NP to target two authentic derived peptides*

3.1 **Overview of the chapter**

In the previous chapter, peptides from digested P-III-NP, pyroE_T1 and hydroxyP_T5, were identified as being the most stable for the subsequent development of quantitative LC-MS methods. Here, synthesized variants (human (*h*) and bovine (*b*)) of these peptides were used to develop MS/MS methods to achieve LODs < 100 pM; thus sufficiently accommodating the analysis of *h*P-III-NP at basal serum concentrations (1-5 ng/mL, 25-125 pM). Matrix effects of serum albumin on P-III-NP peptide analysis are investigated to define the basic requirements of sample preparation. Use of immunoassay (Orion Diagnostica P-IIII-NP assay), as an independent check on P-III-NP concentration throughout sample preparation, was also assessed.

3.2 **Introduction**

In doping control, measurements of P-III-NP have relied on the use of radioactive immunoassays (RIA), RIA-gnost® (CisBio Bioassays) and UniQ® P-III-NP RIA (Orion Diagnostica) and the non-radioactive Advia Centaur® immunoassay platform (Siemens Healthcare Diagnostics Inc.) [59, 150]. In the absence of international reference material for human P-III-NP, these immunoassays use bovine P-III-NP (97 % sequence homology with human) as reference; thus, relying on the cross- reactivity of epitopes for human and bovine material. More recently an enzyme linked immunosorbent assay (ELISA) (CisBio Bioassays) has been developed, however, it is still to be evaluated by anti-doping laboratories.

For each immunoassay, the antibodies and exact epitope location on P-III-NP are unknown. The antibody used in the RIA gnost® assay is based on bovine collagenase digested collagen and has highest affinity (or activity) for the Col 1 domain of P-III-NP. The UniQ® P-III-NP RIA shows activity for the intact 42 kDa protein and immunoreactive molecules with higher molecular weights (e.g. P-III-NP aggregates and

dimers, see Chapter 1 for relevant antigens) [59]. The Centaur P-III-NP assay uses two monoclonal antibodies that bind to two different sites on P-III-NP; however specificity for antigenic form of P-III-NP is unknown [150].

According to immunoassay measurements, circulating P-III-NP has basal serum concentrations between 1-5 ng/mL (25-125 pM) in normal adults [52, 58]. In the absence of international reference material for *h*P-III-NP, immunoassays utilise the cross reactivity of antibodies for bovine and human epitopes [59]. Hence, to develop LC-MS methods targeting human serum P-III-NP, good sensitivity is required to allow for analysis at these small endogenous levels.

3.2.1 ***Mass spectrometry***

Target quantification of proteins by MS can be performed using any tandem MS set up, however, to achieve optimal sensitivity triple quadrupole instruments are often employed [97]. When used for quantitative analysis triple quadrupoles are operated in their selected reaction monitoring (SRM) mode. In SRM mode the precursor analyte as charged ions (e.g. $[hT1+nH]^{n+}$, and $[T5+nH]^{n+}$; where n is any integer except 0) is filtered in the first quadrupole mass analyser (Q1) prior to bombardment with an inert gas to fragment the ion in a collision cell (q). Fragment (product) ions are then selected in the third quadrupole mass analyser (Q3) for detection. Triple quadrupoles are capable of producing scan data, however, due to its mass filtering capabilities more sensitive (greater signal-to-noise, S/N) measurements are achieved in SRM mode. Quadrupoles are considered to be low resolution instruments, resulting from their limited unit *m/z* resolution; as such nominally isobaric background contaminants may be incorporated in the mass filter. To improve selectivity and specificity of MS methods multiple SRMs can be monitored for each analyte.

Peptides (and proteins) need to be present as ions (i.e. possessing a charge) for acceleration and separation through the mass spectrometer hence they are detected as *m/z* ions. Post LC separation, the volatilisation of the accompanying eluent is necessary for the formation of gaseous ions for MS. To achieve this and preserve aa

sequencing information, “soft ionisation” techniques like MALDI (matrix assisted laser desorption ionisation) or ESI (electrospray ionisation) are most frequently applied in protein MS [97]. Normally MALDI produces singly-charged peptides (proteins) ions, unlike ESI, that often produces multiple charged ions that are easily compatible with the limited *m/z* range (< 2000 *m/z*) of triple quadrupole instruments.

In ESI, analytes separated by LC, are ionised during the formation of charged droplets by spraying through a capillary, at high voltage (2-5 kV) with a strong electric field, which is interfaced with the MS. As these charged droplets move towards the MS inlet, the eluent is evaporated to leave droplets containing highly charged analyte ions [151]. To improve sensitivity, micro- and nano-flow LC-MS techniques, which have reduced flow rates (see Table 3-1 below), have been applied to protein MS analyses [152-154]. By using a narrower LC column with reduced flow rates a greater concentration of analyte (relative to solvent) is observed during separation. Also, at these lower flow rates the diameter of the ESI tip can be decreased to result in the formation of smaller droplets, containing less analytes (which reduces clustering) and solvent (leading to rapid desolvation) to improve ionisation efficiency and increase sensitivity.

Table 3-1: Comparison of flow rates and column internal diameters for various HPLC techniques according to Waters Ltd. guidelines [155].

HPLC technique	Typical flow rate	Column i.d.
Conventional	50 µL/min - 2 mL/min	1 mm – 2.1 mm
Micro	10 – 100 µL/min	500 µm – 1 mm
Capillary	3 – 10 µL/min	150 – 300 µm
Nano	100 nL/min – 1 µL/min	75 – 100 µm

3.2.1.1 Micro-flow LC-MS

Miniaturisation of the LC system to facilitate micro flow rates (10-100 µL) has resulted in the improved sensitivity of many proteins (peptides) [152, 156, 157]. However, reducing the internal diameter (i.d.) of the LC column limits the sample injection volumes that can counterbalance the theoretically improved sensitivity. To avoid peak broadening and maintain the separation of micro-LC columns, sample volumes < 30 % of the column peak capacity should be injected [153], dead volume is also to be

minimised by reducing the size of connecting tubing and the diameter of the ESI needle [158].

3.2.1.2 Nano-flow LC-MS

Further reduction in the sample injection volume is needed for direct injection onto nano-LC columns; for which this compromise does not facilitate an increase in sensitivity. To overcome this, nano-systems use a column switching mechanism, where the sample (typically 5 µL) is loaded onto a short wider i.d. "trapping column" (18 µm x 20 mm) at a relatively high flow rate (using a mobile phase composition that will retain the analyte) before loading the analyte by mass onto the analytical column (75 µm x 150 mm). See Figure 3-2 for systematic diagram of valve switching mechanism that dictates LC flow. Once injected onto the nano-analytical column, separation is achieved using nano-LC flow rates. The i.d. of the ESI emitter of the nano- is also reduced (typical range 5-30 µm), and is normally made of fused silica or metal. With these i.d.'s much smaller droplet ions are formed at the end of the emitter, hence, nano-emitters are conveniently placed closer to the inlet of the MS, allowing for an in-line rather than orthogonal connection, which maximises the detection of the vaporised analytes, as shown in Figure 3-1.

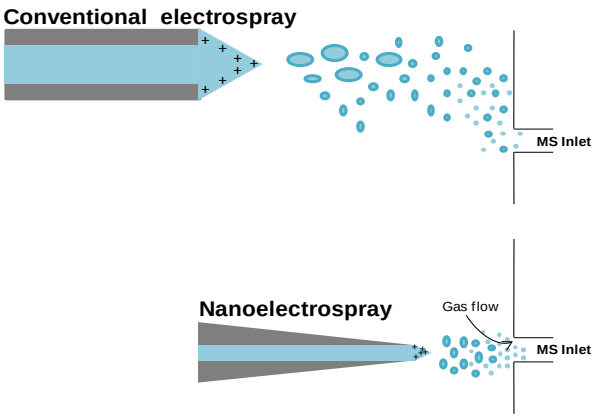


Figure 3.2-1: Illustration showing droplet formation at the ESI tip used in conventional and nano-electrospray MS. Difference in size and position of the nanospray emitter tip to the MS inlet, allows for more analyte ions to enter to the MS, theoretically increasing sensitivity.

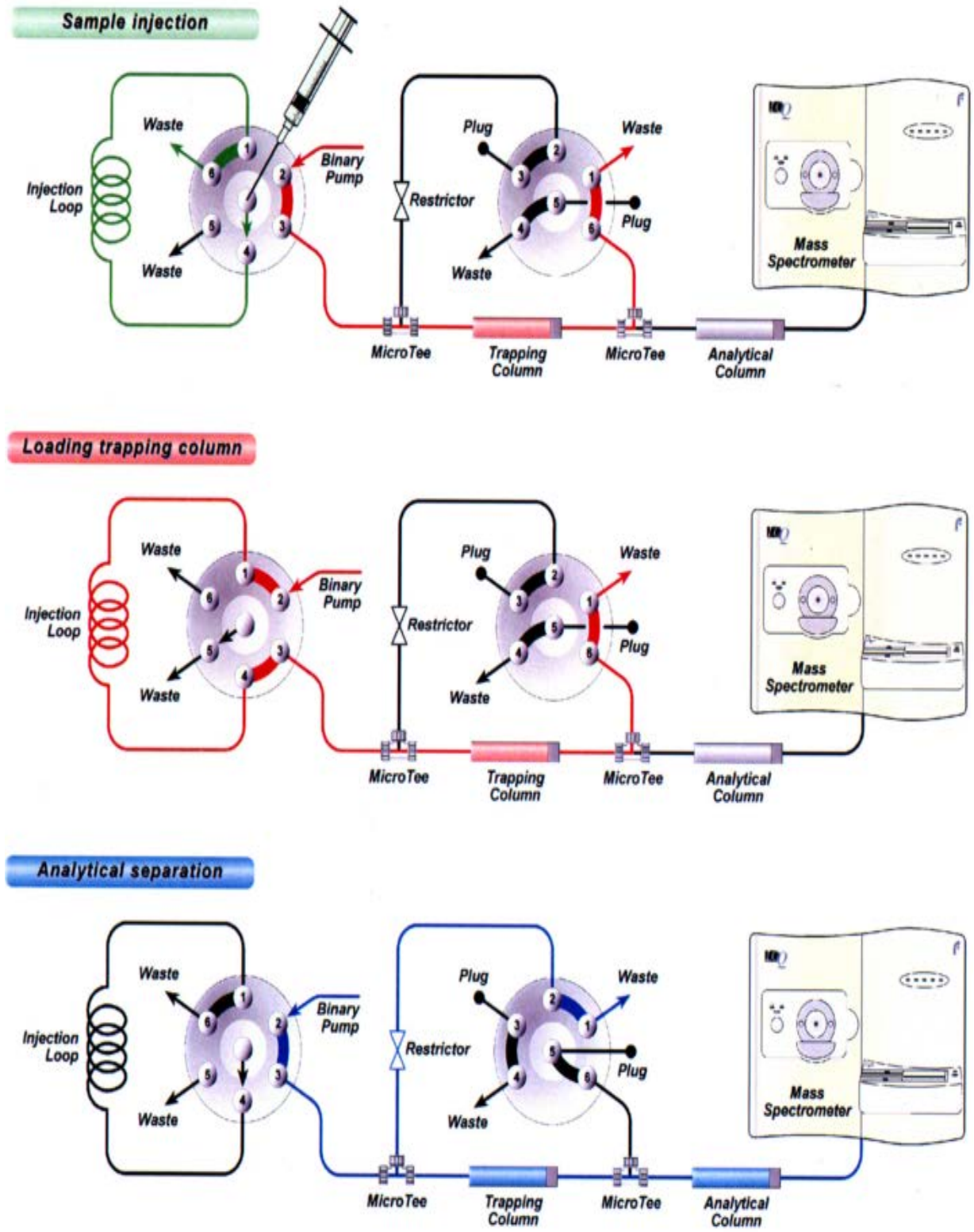


Figure 3.2-2: Schematic of the sample injection, trapping and separation process used for nano-flow LC-MS, showing the valve position used during column switching. During sample loading an injection flow is directed to the trapping column only. Post trapping, the flow is directed to the analytical column through the trapping column; where the analytes are transferred and subsequently separated using gradient elution (image taken from [153] with permission).

3.3 **Materials and Methods**

The source of materials where unmentioned can be found in section 2.3. Authentic human and bovine P-III-NP peptides, *pyro*E_hT1 (human, *h*T1), *pyro*E_bT1 (bovine, *b*T1) and *hydroxy*P_T5 (T5), were synthesized by Fisher Scientific (Offenbach, Germany). The aa sequence of each peptide (500 fmoles) was verified by LC-MS using the HRMS methods described in section 2.3.3.2.2.

3.3.1 ***Mass spectrometry***

Similar to the development of HRMS methods (section 2.3.3.2.1), Substance P (371 nM) was used to define initial source parameters for analysing P-III-NP peptides on a Waters Xevo® TQ-S MS triple quadrupole system. A flow of 10 µL/min was used for conventional ESI methods whilst flow rates of 8 µL/min and 0.4 µL/min were used for micro and nano ESI respectively. Analysis was carried out in positive ion mode using N_2 as the auxiliary gas. Source parameters were retained for conventional ESI methods, however, for micro- and nano-ESI further optimisation, with synthesized peptides (500 nM), at their respective flow rates was required to establish the parameters detailed in Table 3-2 below.

Table 3-2: Tune file parameters for each ESI set up on a Waters Xevo® TQS MS

Instrument Parameter	Conventional- ESI	Micro- ESI	Nano- ESI
Capillary (kV)	2.78	2.22	1.75
Cone (V)	26.00	22.00	45.00
Source Offset (V)	27.0	27.0	30.0
Source Temperature (˚C)	150	150	83
Desolvation Temperature (˚C)	350	350	-
Cone Gas Flow (L/Hr)	150	145	-
Desolvation Gas Flow (L/h)	700	600	-
Collision Gas Flow (mL/min)	0.34	0.17	0.15
Nebuliser Gas Flow (Bar)	7.00	7.00	7.00
LM 1 Resolution	2.3	2.3	3.2
HM 1 Resolution	13.7	13.7	11.4
Ion energy 1	1.2	1.2	0.8
LM 2 Resolution	2.8	2.8	2.8
HM2 Resolution	14.9	14.9	14.9
Ion energy 2	0.6	0.6	0.6
MS mode collision energy	4.00	4.00	10.00

3.3.1.1 Selected reaction monitoring

For each peptide (~500 nM), a full scan spectrum was collected with each ESI source by LC-MS to select appropriate precursors for SRM experiments. Using the most abundant precursor for each peptide, collision energy (CE) profiles between 5 - 50 eV were established from product ion scan data to select the MS-MS transitions to monitor in SRM using conventional ESI for *h*T1 and T5 and nano-ESI for *b*T1, listed in Table 3-3 below.

Table 3-3: Collision energies for SRM (selected reaction monitoring) transitions acquired for P-III-NP peptides by conventional-, micro- and nano-flow ESI MS using a Waters Xevo® TQ-S.

P-III-NP Peptide	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (eV)
***h*T1**	692	768	10
	1038	240	50
	1038	440	40
***b*T1**	688	539	20
	688	761	25
	1031	440	45
T5	534	363	20
	534	448	13
	534	628	27

3.3.2 *Liquid chromatography*

On a Waters Acquity® UPLC® coupled to a Xevo® TQ-S, the simple LC gradient conditions described in section 2.3.3.1 were used to approximate %B (0.3 % formic acid in acetonitrile) required to elute hT1 (and *b*T1) and T5 from an Acquity® UPLC® BEH 130 Å (C18, 1.7 µm, 2.1 x 50 mm) column. Dead time was established using the elution of unretained acetone (10 µL). As a basic model this was used to develop conventional-, micro- and nano-flow LC methods to enable at least 12 points per peak for each MS-MS transition acquired in SRM; whilst minimising any ion suppression that could be contributed from matrix (especially tryptic digest products of HSA).

3.3.2.1 Conventional LC

Conventional-flow LC-MS was performed on a Waters Acquity® UPLC® coupled to a Waters Xevo® TQ-S MS interfaced with a conventional ESI ion source. Gradient LC separation of each sample (10 µL) was achieved using an Acquity® UPLC® BEH 130 Å

(C18, 1.7 µm, 2.1 x 100 mm) column at 40 °C, with mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile and a flow rate of 300 µL/min. Mobile phase B was held at 5 % for 3 min before increasing to 20 % in 17 min, then to 90 % in 1 min and held for a further minute. Mobile Phase B was then immediately returned to 5 % and held for 3 min to re-equilibrate the column.

3.3.2.2 Micro LC

Micro-flow LC-MS was performed on a Waters nanoAcquity® UPLC® coupled to a Waters Xevo® TQ-S MS interfaced with a narrow gauge ESI needle (60 µm i.d.) on a conventional ion source. A low dead volume fused silica capillary (25 µm i.d.) was used from the outlet of the column to the inlet of the ESI needle. Gradient LC separation of each sample (1 µL) was achieved on an Acquity® M - Class BEH 130 Å (C18, 1.7 µm, 300 µm x 150 mm) column with mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile and a flow rate of 8 µL/min. Mobile phase B was increased from 5 % to 60 % over 20 min, then to 90 % in 1.5 min and held for 30 sec, after which it was immediately returned to 5 % and held for 3 min to re-equilibrate the column.

3.3.2.3 Nano LC

Nano-flow LC-MS was performed on a Waters nanoAcquity® UPLC® coupled to a Waters Xevo® TQ-S MS interfaced with a nano-ESI emitter (30 µm i.d.) using the Waters nano-spray source. Gradient LC separation was achieved using mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile. Samples (5 µL) were trapped on an Acquity® UPLC® Symmetry C18 nanoAcquity® 10K 2GV/M trapping column (100 Å, 5 µm, 180 µm x 20 mm) for 5 min at flow 10 µL/min with 1 % mobile phase B. The trapped analytes were then loaded onto an Acquity® UPLC® M-Class Peptide BEH C18 column (130 Å, 1.7 µm, 75 µm x 150 mm) and separated by gradient elution. With a flow rate of 0.4 µL/min, mobile phase B was held at 5 % for 5 min, then increased to 9 % for 25 min and then 95 % for 2 min and held for 1 min. Mobile Phase

B was then immediately returned to 5 % and held for 27 min to re-equilibrate the column.

3.3.3 *Limit of detection and carryover*

Using pure peptides for *h*T1 and T5, peptide standards (500 nM) were prepared as described in section 2.3.2.1. In parallel, dilution standards (5, 10, 20, 50, 100, 200, 500, 1,000 pM) were prepared from a sub-stock (1,000 pM) in mobile phase (0.3 % formic acid, 5 % acetonitrile in water) and analysed using the developed conventional-, micro- and nano- LC-MS methods in increasing concentration order (i.e. 5 pM to 10,000 pM). LOD was established where the S/N of the peptide was < 3:1. Mobile phase was injected before and after each injection in order to establish LOD of the peptides and assess carryover associated with LC method.

3.3.4 *Effect of matrix on P-III-NP LC-MS analysis*

Bovine P-III-NP 100 µg/mL (100 µL) and 10 µg/mL (100 µL) were separately spiked into a sample of HSA (100 mg/mL, 100 µL) and digested with trypsin (10 µg) using the protocol described in section 2.3.4. Once digested, samples were analysed by conventional LC-MS.

3.3.4.1 HSA cut-off for P-III-NP peptide LC-MS analysis

Duplicate HSA standards (0.1, 0.5, 1, 5, 10, 25, 50, and 100 µg) were digested with trypsin (2 µg) using the protocol described in section 2.3.4. One of each standard was reconstituted in formic acid (0.3 %) and acetonitrile (5 %) solution (50 µL) containing either P-III-NP *h*T1 and T5 peptides at 100 pM or at 500 pM; before transferring to an autosampler vial. The samples containing 500 pM *h*T1 and T5 were analysed by conventional-flow LC-MS whilst those containing 100 pM *h*T1 and T5 were analysed by nano-flow LC-MS to determine the effect of HSA derived tryptic peptides on P-III-NP peptides analysis.

3.3.5 ***Concentration of P-III-NP and synthesised peptides - Immunoassay***

The concentration of intact P-III-NP (~50 ng/mL) and equivalent concentrations of desalted P-III-NP fragments (single stranded monomeric P-III-NP, carbamidomethylated *h*T1 (and *b*T1) and T5) was measured using the Orion UniQ PIIINP radioactive immunoassay kit (Orion Diagnostica, Espoo, Finland) as follows.

Whilst reagents, calibrators (0- 50 ng/mL), control (human serum) and samples were brought to room temperature (18-25 ˚C), test tubes were labelled for non-specific binding (NSB), calibrators, control, and samples. In duplicate, calibrators (200 µL), samples (100 µL) control (100 µL), and NSB (100 µL of any sample) were pipetted into appropriate tubes. To make up the volume of the sample, control and NSB tubes, HSA (0.1 %, 100 µL) was added. Tracer (200 µL) was added to all tubes. Antiserum (200 µL) was added to all tubes except the NSB, to which distilled water (200 µL) was added. All tubes were vortexed then covered with paraffin film and incubated for 2 h at 37 ˚C. After mixing thoroughly by gently inverting, procollagen separation reagent (500 µL) was added to all tubes before vortexing and incubating for 30 min at room temperature. Tubes were centrifuged (2000 g) for 15 min at 4 °C before carefully decanting the supernatant (except the totals) by simultaneous inversion taking care not to disturb the precipitate. Using a gamma counter each tube was counted for at least 1 min or until 10,000 counts/tube was accumulated. Using the mean count of calibrators, samples and controls the %B/B0 was calculated using Equation 3-1. On semi-log graph paper with %B/B0 values on the ordinate and P-III-NP concentration on the abscissa a calibration curve was drawn, from which the P-III-NP concentrations of samples and controls were determined.

Equation 3-1:Calculation of %B/B_0 for Orion P-III-NP RIA

$$\%\frac{B}{B_0} = \frac{calibrator\ or\ sample\ count - NSB}{calibrator\ 0\ count - NSB} \times 100$$

3.4 **Results and Discussion**

3.4.1 ***LC- MS optimisation***

3.4.1.1 LC gradient

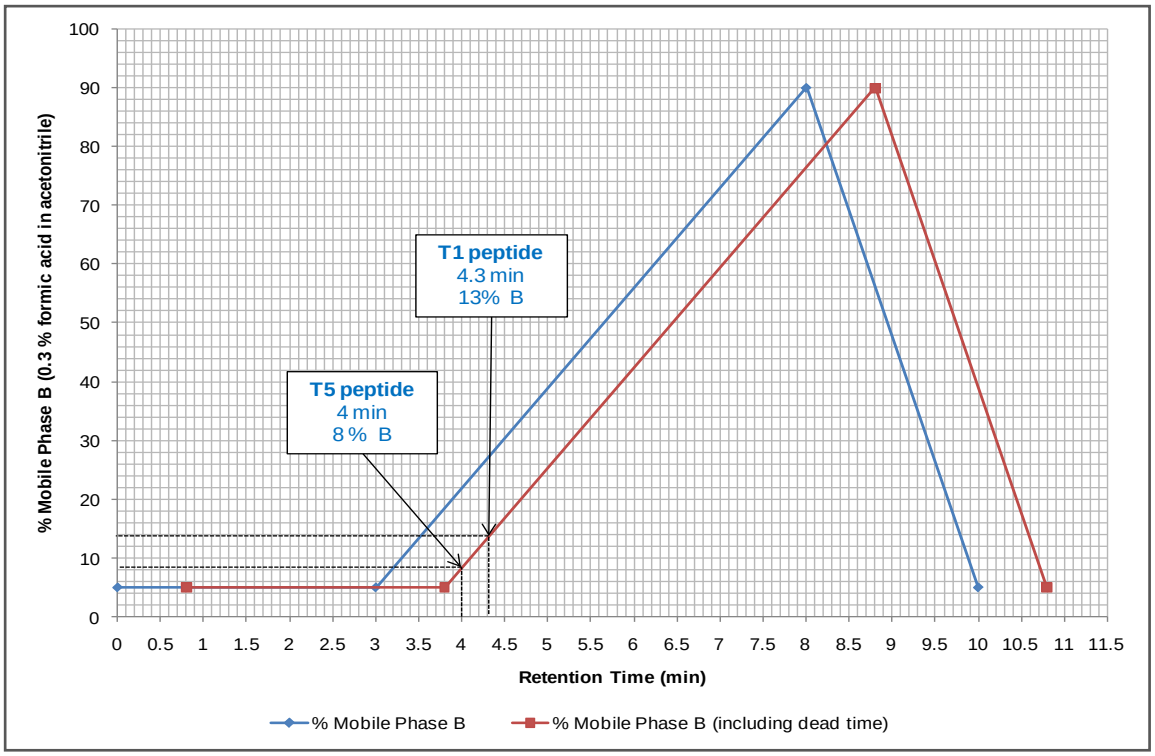


Figure 3.4-1: Gradient elution of P-III-NP pyroE_T1 (T1) and hydroxyP_T5 (T5) peptides using simple LC conditions (described in section 2.3.3.1) on an Acquity® BEH C-18 column at 40 ˚C. The elution mobile phase composition (as % B, acetonitrile) of each peptide is shown.

It is expected that T5, the less hydrophobic peptide would elute first during LC separation on a reversed phase column. Having only one hydrophobic aa (I-8) and several hydroxylated prolines (P-3, 6 and 9) resulting from PTM, T5 has limited interaction with the non-polar C-18 stationary phase of the column. Originating from the globular Col1 domain of P-III-NP, T1 contains more hydrophobic aa (A-4, V-5, L-12 and Y-16), thus increased column interaction is expected (and is observed) by this peptide.

Despite a 5 % eluent composition difference between the P-III-NP peptides, using the simple LC conditions described in section2.3.3.1, the observed retention times of these peptides are similar (T5- 4 min and T1- 4.3 min). Thus, to maximise separation (and resolve P-III-NP peptides from potentially similar matrix-generated peptides), the slope

within the peptide eluting region (% B 8 →12) is decreased in the developed LC-MS methods.

For conventional-flow LC increasing the concentration of mobile phase B at a rate of 0.88 % per min (flow 300 µL/min) will obtain a t_R difference of 5 min between T5 and T1 peptides. Optimal t_R difference (~15 min) between peptides was achieved with nano-flow LC, where the concentration of mobile phase B was increased at a rate of 0.16 % and then 44 % per min (flow 0.4 µL/min) in the peptide eluting region. For micro-LC, however, although utilising a nano-LC pump capable of very low flow rates, a relatively fast flow (8 µL/min) and gradient increase for mobile phase B (2.75 % per min) was used to improve the peak shape of T1. Thus a t_R difference of only 1.2 min was observed between peptides for the micro-LC method. For all developed LC methods, baseline resolution of T5 and T1 was achieved.

3.4.1.2 SRM - collision energies

3.4.1.2.1 *HydroxyP*_T5 peptide ($GDP_{+OH}GPP_{+OH}GIP_{+OH}R$)

In full scan, the dominant T5 precursor, $[M+2H]^{2+}$ (*m/z* 534) was observed. With application of argon and various collision energies, fragmentation of the peptide was achieved to produce sufficient diagnostic transitions for SRM methods (see Figure 3-4 below). Within the CE profile the y_9^{2+} peptide (*m/z* 448) is the most abundant. This peptide is generated at a relatively low CE, (optimal 13 eV) as a result of the fragility of the $^2D\text{-}^3P_{+OH}$ bond at the cleavage point within T5. Further increase in CE results in the reduction of this peptide, which is truncated through secondary fragmentation to produce smaller peptide fragments y_7^{2+} (*m/z* 363) and y_6 (*m/z* 628). Typical to the product scan pattern, generated by MS-HRAMS (Chapter 2), the dominant peptide fragments terminate with a P (or P_{+OH}) residue. These data suggest that the natural twist of P influences the MS peptide fragmentation pattern. It is interesting to note that peptide fragments (y_9^{2+} and y_7^{2+}), which contain the $^5P\text{-}^6P_{+OH}$ bond, retains both of the

available protons. At CE 27 eV, significant loss of peptide fragments was observed, thus further fragmentation was not investigated.

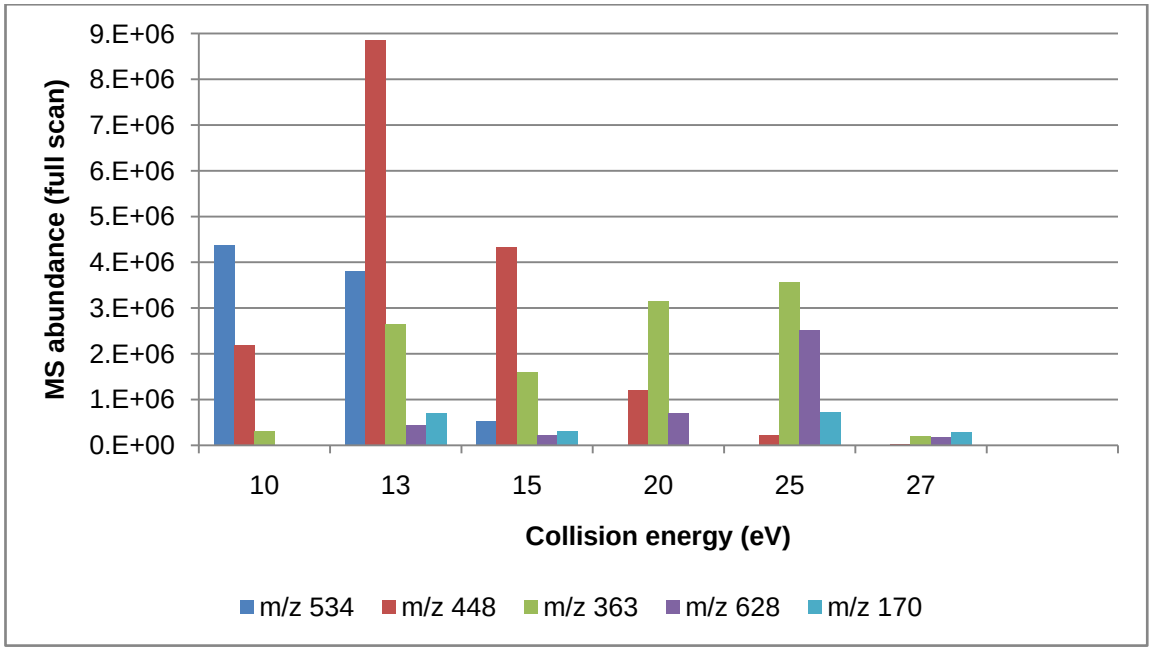


Figure 3.4-2: Collision energy profile (10-27 eV) for P-III-NP T5 peptide (500 pM) obtained from product scan spectra of $[M+2H]^{2+}$ precursor (*m/z* 534 ±1 amu) using conventional ESI-MS conditions.

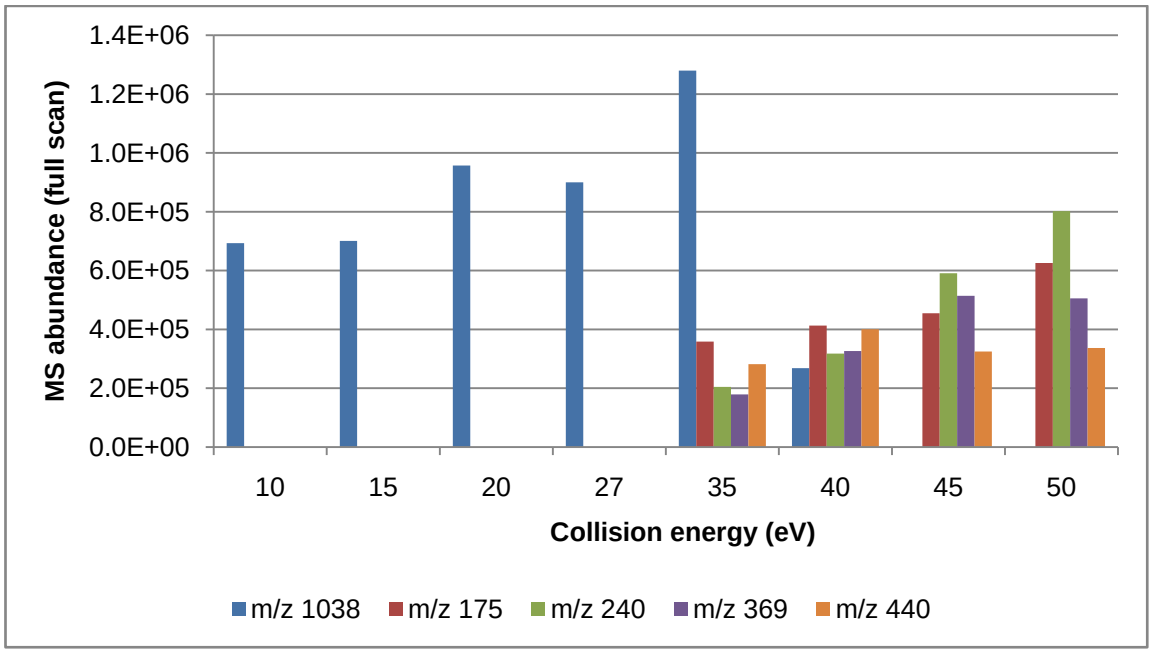


Figure 3.4-3: Collision energy profile (10-50 eV) for P-III-NP *h*T1 peptide (500 pM) obtained from product scan spectra of $[M+2H]^{2+}$ precursor (*m/z* 1038 ±1 amu) using conventional ESI-MS conditions.

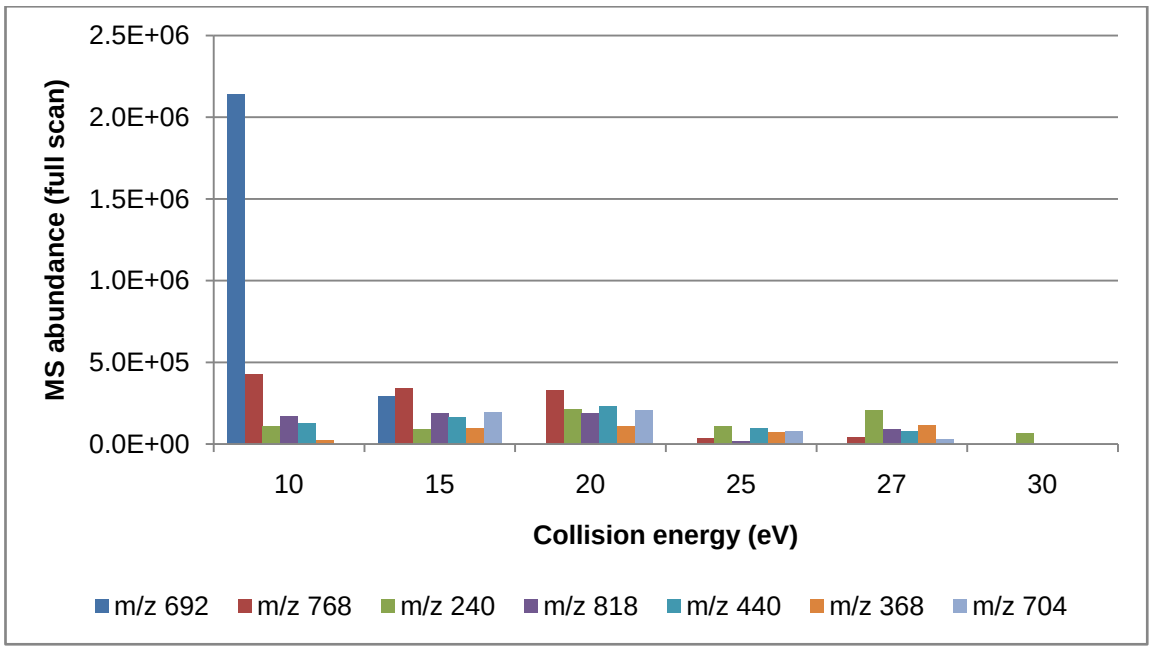


Figure 3.4-4: Collision energy profile (10-30 eV) for P-III-NP *h*T1 peptide (500 pM) obtained from product scan spectra of $[M+3H]^{3+}$ precursor (*m/z* 692 ±1 amu) using conventional ESI-MS conditions.

3.4.1.2.2 PyroE_*h*T1 (Q_{-NH3}QEAV*E*GGCSHLGQSYADR)

For *h*T1, peptide fragments generated from the $[M+2H]^{2+}$ (*m/z* 1038, Figure 3-5) and $[M+3H]^{3+}$ (*m/z* 692, Figure 3-6) precursors were investigated. Fragmentation of the $[M+2H]^{2+}$ precursor requires significantly higher collision energies (≥ 35 eV), when compared to T5. As a result, smaller less diagnostic peptide fragments were observed in the CE profile shown in Figure 3-5 above (e.g. y_1^+ (*m/z* 175), b_2^+ (*m/z* 240), b_3^+ (*m/z* 369) and b_4^+ (*m/z* 440)). For the $[M+3H]^{3+}$ precursor, however, the presence of an additional "mobile proton" results in a decrease in the CE necessary for fragmentation, with complete fragmentation occurring when a CE ≥ 30 eV is applied. Similar to the $[M+2H]^{2+}$ precursor, small peptides are generated, however, diagnostic peptide fragments y_{14}^{2+} (*m/z* 768) and y_{13}^{2+} (*m/z* 704) are also observed. Although significantly lower in abundance than the smaller peptide fragments generated by $[M+2H]^{2+}$ precursor (*m/z* 1038), the $[M+3H]^{3+}$ (*m/z* 692)→ y_{14}^{2+} (*m/z* 768) transition, is monitored in SRM in the developed methods (see section 3.3.1.1 for transitions).

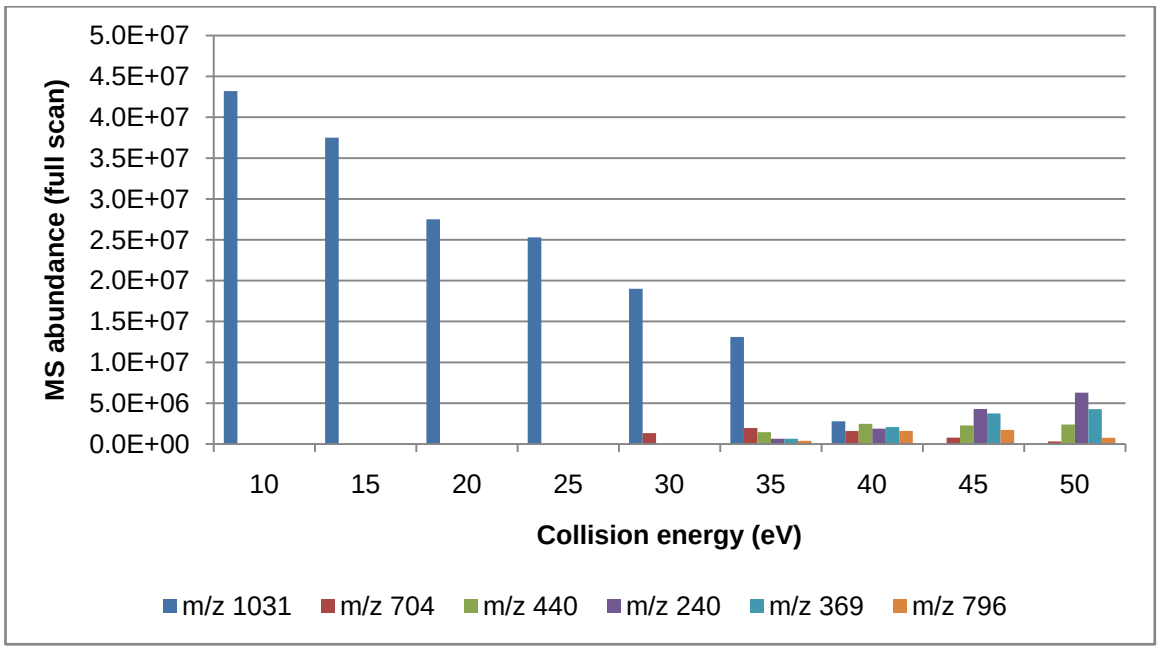


Figure 3.4-5: Collision energy profile (10-50 eV) for P-III-NP *b*T1 peptide (500 pM) obtained from product scan spectra of $[M+2H]^{2+}$ (*m/z* 1031 ±1 amu) using nano ESI-MS conditions.

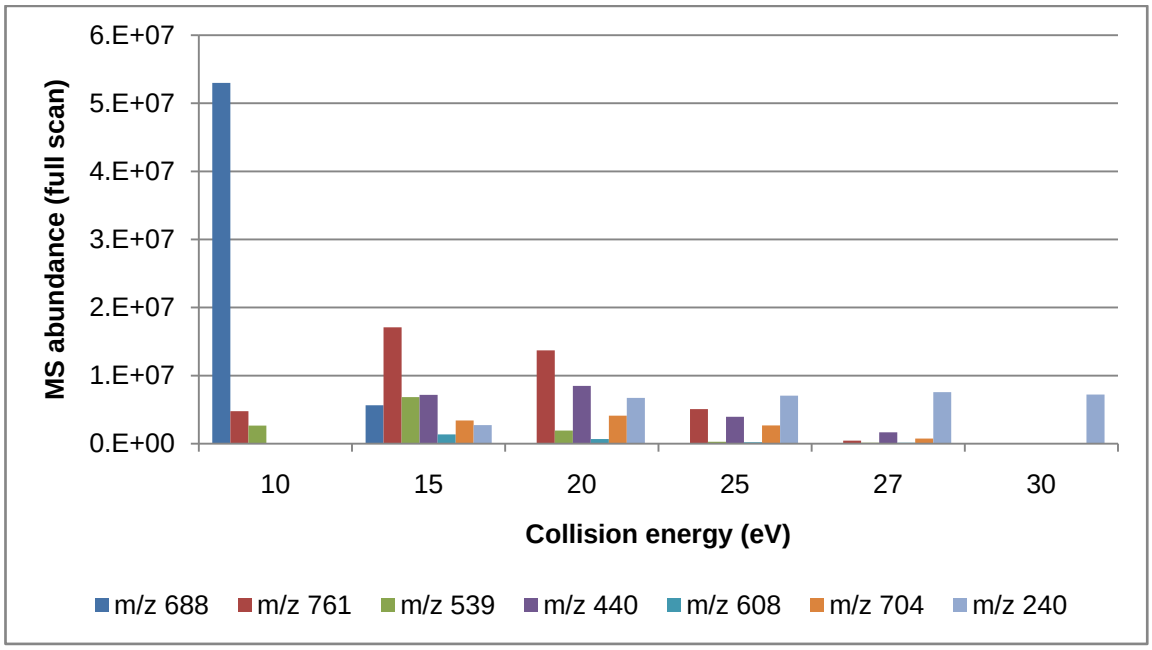


Figure 3.4-6: Collision energy profile (10-30 eV) for P-III-NP *b*T1 peptide (500 pM) obtained from product scan spectra of $[M+3H]^{3+}$ (*m/z* 688 ±1 amu) using nano ESI-MS conditions.

3.4.1.2.3 PyroE_*b*T1 (Q_{-NH3}QEAV*D*GGCSHLGQSYADR)

Similar to the CE profile produced by the $[M+2H]^{2+}$ *h*T1 precursor, the bovine $[M+2H]^{2+}$ peptide (*b*T1, *m/z* 1031, Figure 3-7) requires a high CE for fragmentation. Diagnostic ions y_7^+ (*m/z* 796) and y_{13}^{2+} (*m/z* 704) are observed in the peptide profile along with the less diagnostic y_1^+ (*m/z* 175), b_2^+ (*m/z* 240), b_3^+ (*m/z* 369) and b_4^+ (*m/z* 440) ions. The appearance of these ions may be attributed to the increased sensitivity provided by the nano-ESI source used to generate this profile, unlike the conventional-ESI used for the *h*T1 and T5 peptides. However, the abundance of the peptides produced by $[M+2H]^{2+}$ (*m/z* 1031)→ y_7^+ (*m/z* 704) and $[M+2H]^{2+}$ (*m/z* 1031)→ y_{13}^{2+} (*m/z* 704) is significantly lower than that generated from the $[M+3H]^{3+}$ precursor (*m/z* 688), hence these transitions were excluded from SRM methods. For the $[M+3H]^{3+}$ precursor, like *h*T1, peptide fragments are generated at lower CE with complete fragmentation occurring at CE ≥ 30 eV, with significant loss in abundance of the more diagnostic peptide fragments at CE 25 eV.

3.4.2 ***Limit of detection, carryover and linearity***

In Chapter 2, the sensitivity achieved from LC-MS/HRMS methods for P-III-NP peptides was insufficient for analysis of serum P-III-NP at endogenous levels. To increase sensitivity, methods have been developed on a highly sensitive Waters Xevo® TQ-S triple quadrupole mass spectrometer utilising its mass filtering capabilities in selected reaction monitoring mode. Three ESI set-ups have been investigated on this instrument and are compared here. For conventional- and nano-flow LC methods, the injection volume (10 and 5 µL, respectively) was assigned based on the capacity of the sample loop on the system. For micro-flow LC, however, volumes ≥ 2 µL resulted in volume overload, so, an injection volume of 1 µL was selected for this method. For each peptide LOD was assigned at the lowest standard where all monitored SRMs had a signal to noise ratio > 3:1. For some set-ups (e.g. nano-flow LC-MS, see Figure 3-9 below), almost complete elimination of noise was achieved; hence, LODs were determined at the limit of the mass signal detected by the MS.

3.4.2.1 PyroE_*h*T1 (Q_{-NH3}QEAV***E***GGCSHLGQSYADR)

For conventional-, micro- and nano-flow LC-MS the approximate LOD for *h*T1 was determined to be 100 pM (10 µL, 1 fmole), 100 pM (1µL, 100 amoles) and 5 pM (5 µL, 25 amoles), respectively (see Figure 3-9). These LODs were obtained in the absence of matrix peptides, hence they represent instrumental LODs. The observed improvement in sensitivity for peptides is associated with decreased flow of the LC (and essentially the decrease in the volume of solvent accompanying the analyte), which improves ionization efficiency of the ESI. Despite the LOD (mass) of micro-flow LC-MS, the linearity of standards in producing a calibration line is very poor ($R^2 < 0.9$) when compared with that obtained by conventional- and nano-flow methods for the same standards. For conventional- and micro-flow LC-MS methods, the area (and height) counts observed for *h*T1 are low (even at 1000 pM), further highlighting the inefficiency of ionisation of this peptide. Despite this, unlike the LOD achieved by LC-MS/HRMS (7 nM, see section 2.4.2.3), the established conventional-, micro- and nano-flow LODs are suitable for analysis of *h*T1 originating from serum P-III-NP at its highest expected basal level (5 ng/mL, 125 pM); with nano-flow LC-MS capable of analysis at the lower limit (1 ng/mL, 25 pM).

Considering that *h*T1 and T5 peptides were analysed together using parallel dilution from the same sub-stock solution (1 nM *h*T1 and T5), the linearity of *h*T1 (R^2 between 0.873-0.992) was noticeably less than that achieved for T5 (R^2 between 0.991-0.999, see section 3.4.2.2 below). It could be argued that the observed variance is due to a lack of homogeneity of *h*T1 in the stock, originating from the presence of oxidised and reduced forms of the *h*T1 due to incomplete carbamidomethylation during derivatisation. To investigate this, the carbamidomethylated *h*T1 stock (500 nM, 2 µL) was analysed by LC-MS/HRMS (see section 2.3.3 for details). A very small amount of reduced *h*T1 (0.24 % of carbamidomethylated *h*T1) at *m/z* 1009.4373 (± 5 ppm) was observed (see Figure 3-13). No oxidised *h*T1 (*m/z* 1008.9331) was observed, however, this could be due to the limited sensitivity of the Orbitrap instrument compared with the

triple quadrupole. Regardless, at these negligible concentrations, the presence of reduced and oxidised *h*T1 should not affect the linearity of the peptide.

Carryover was not observed for the *h*T1 peptide, by any of the developed LC-MS methods. Hence, this could not affect the observed linearity. To improve the linearity, normalisation of the data with the addition of an internal heavy-labelled (i.e. C-13/ N-15) reference material with similar ionisation properties should be investigated.

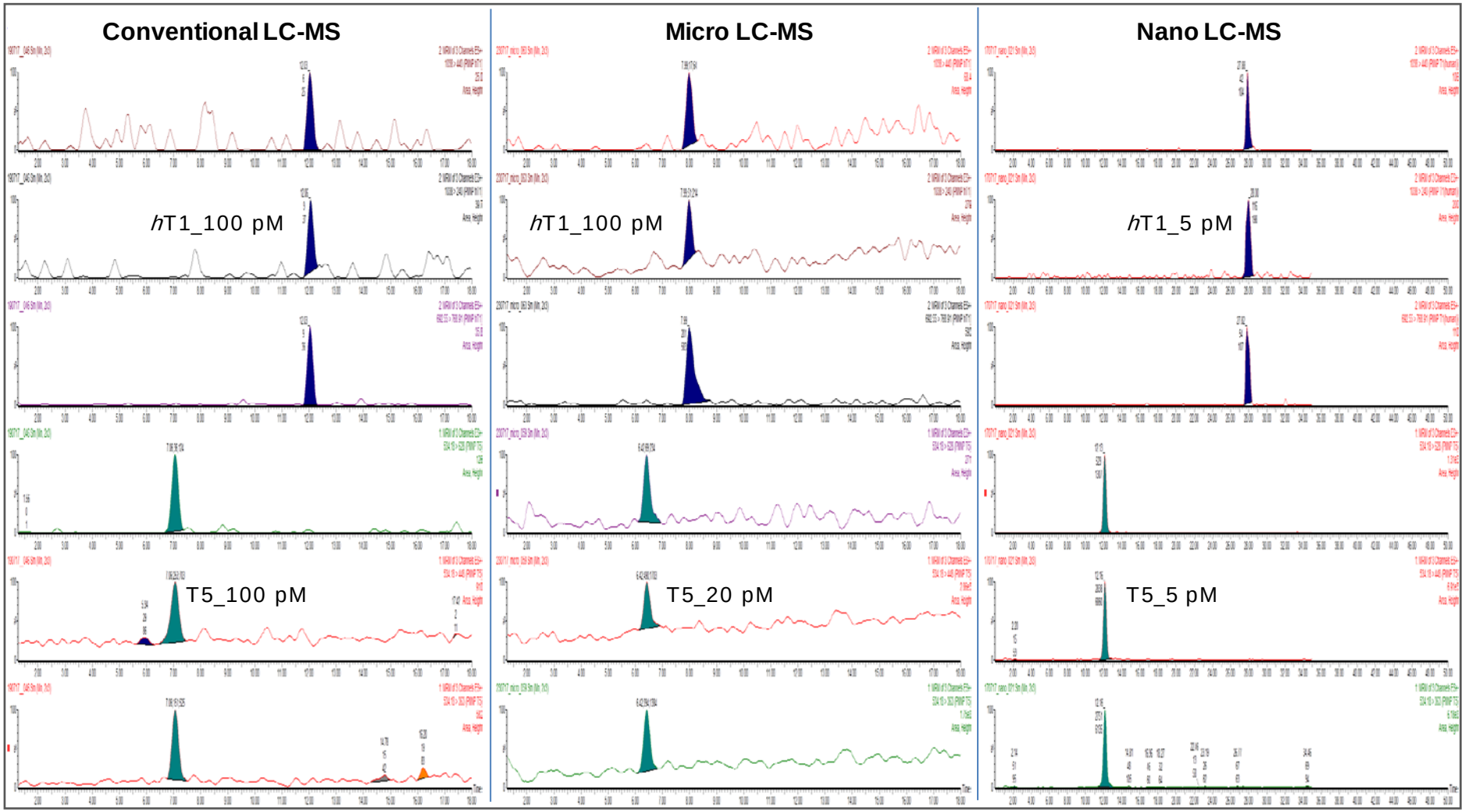
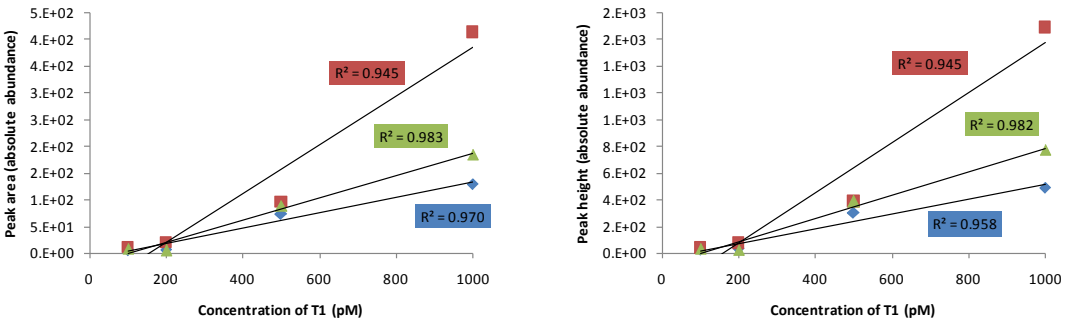
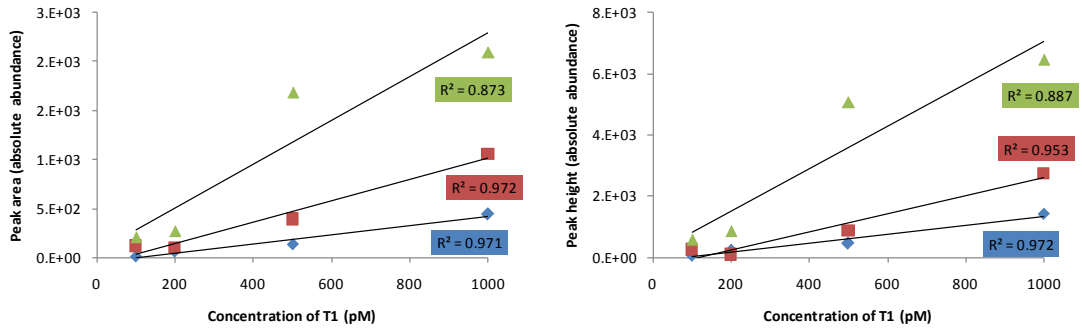


Figure 3.4-7: Chromatogram of P-III-NP peptides (*h*T1 and T5) near their LODs for conventional-, micro- and nano-flow LC-MS using optimised SRM methods.



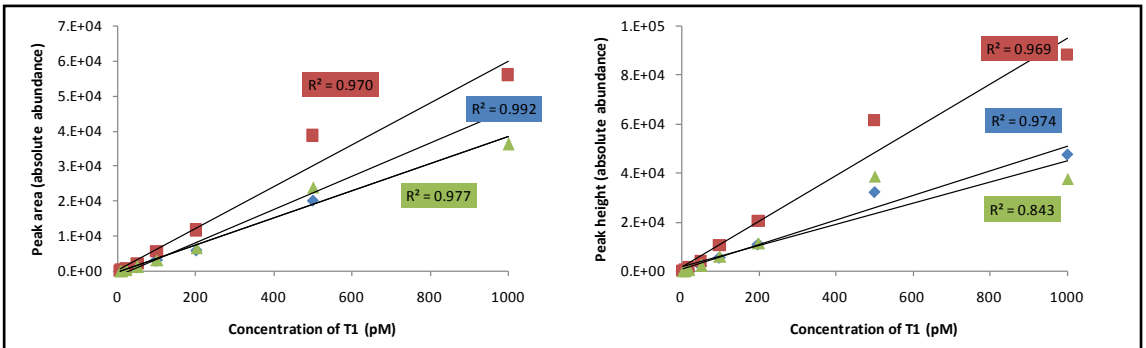
Concentration of T1 (pM)	Peak area (conventional-flow LC-MS)			Peak height (conventional-flow LC-MS)		
	m/z 1038 -> 440	m/z 1038 -> 240	m/z 692 -> 768	m/z 1038 -> 440	m/z 1038 -> 240	m/z 692 -> 768
100	6	11	9	25	40	36
200	7	20	6	31	77	27
500	74	96	90	303	389	388
1000	129	414	186	493	1694	777

Figure 3.4-8: Linearity of synthesised P-III-NP *h*T1 peptide standards analysed by conventional-flow LC-MS using absolute peak area and height.



Concentration of T1 (pM)	Peak area micro-flow LC-MS)			Peak height (micro-flow LC-MS)		
	m/z 1038 -> 440	*m/z* 1038 -> 240	*m/z* 692 -> 768	*m/z* 1038 -> 440	*m/z* 1038 -> 240	*m/z* 692 -> 768
100	17	120	212	64	267	591
200	64	103	275	219	103	884
500	141	392	1684	466	875	5079
1000	442	1054	2099	1428	2743	6482

Figure 3.4-9: Linearity of synthesised P-III-NP *h*T1 peptide standards analysed by micro-flow LC-MS using absolute peak area and height.



Concentration of T1 (pM)	Peak area (nano-flow LC-MS)			Peak height (nano-flow LC-MS)		
	m/z 1038 -> 440	*m/z* 1038 -> 240	*m/z* 692 -> 768	*m/z* 1038 -> 440	*m/z* 1038 -> 240	*m/z* 692 -> 768
5	45	69	32	106	198	109
10	85	223	78	217	430	162
20	351	696	263	651	1411	677
50	1012	2144	1200	2004	4172	2353
100	2923	5627	3114	5567	10542	5971
200	6047	11682	6523	10934	20387	11457
500	20099	38783	23871	32305	61559	38583
1000	47566	56061	36289	47566	88412	37728

Figure 3.4-10: Linearity of synthesised P-III-NP *h*T1 peptide standards analysed by nano-flow LC-MS using absolute peak area and height.

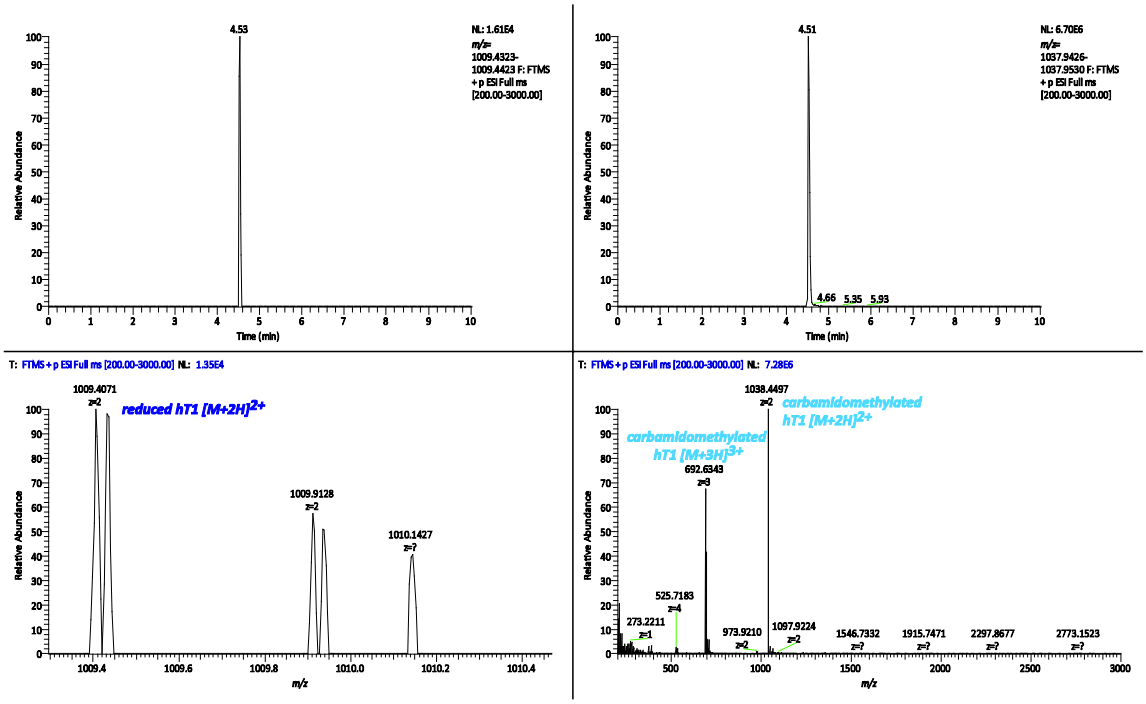
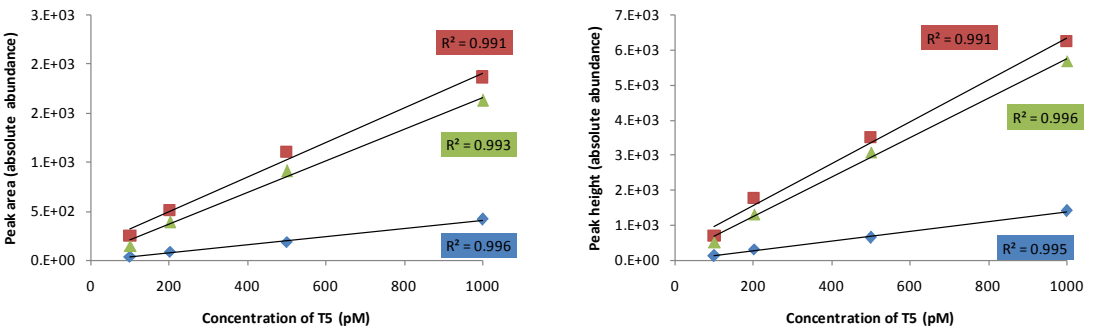


Figure 3.4-11: Extracted ion chromatogram at *m/z* 1009.4373 ±5 ppm (reduced *h*T1 $[M+2H]^{2+}$ peptide) and *m/z* 1037.9478 ± 5ppm (carbamidomethylated *h*T1 $[M+2H]^{2+}$ peptide) and corresponding extracted HRMS spectrum.

3.4.2.2 HydroxyP_T5 ($GDP_{+OH}GPP_{+OH}GIP_{+OH}R$)

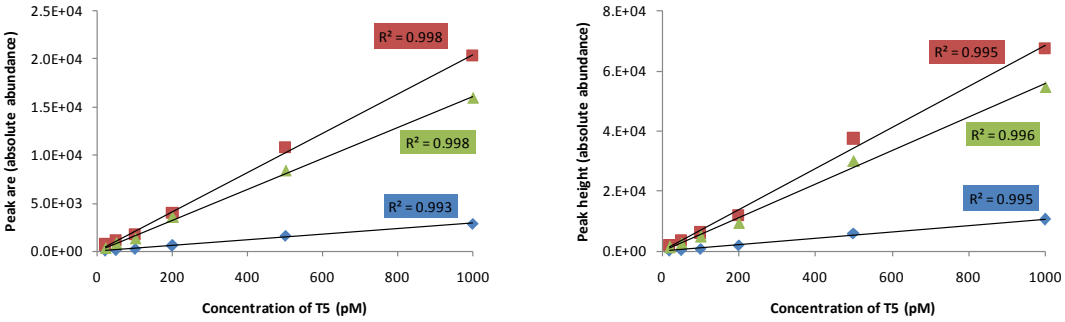
For conventional-, micro- and nano-flow LC-MS the approximate instrumental LOD for T5 was 100 pM (10 µL, 1 fmole), 20 pM (1 µL, 20 amoles) and 5 pM (25 amoles), respectively (see Figure 3-9). Like *h*T1, with decreased flow a significant improvement in LOD was observed from conventional- to micro-flow methods. However, the LODs observed for nano- and micro-flow suggests that 20 amoles is the true detection limit for T5 on the Xevo® TQ-S triple quadrupole instrument. This suggests that optimum droplet size is achieved with micro-ESI using the narrow gauge needle (i.d. 60 µm).

Despite the absence of an "internal standard" to normalise the data set, good linearity of the T5 standards was achieved. Compared to *h*T1, significantly greater sensitivity is observed for the T5 peptide, hence, this should be the preferred peptide for P-III-NP quantification. With the established LODs, serum P-III-NP at basal levels is possible, especially when micro- and nano-flow methods are used. Negligible carryover (< 0.005 %) is observed for T5 peptide (1000 pM) by nano-LC-MS.



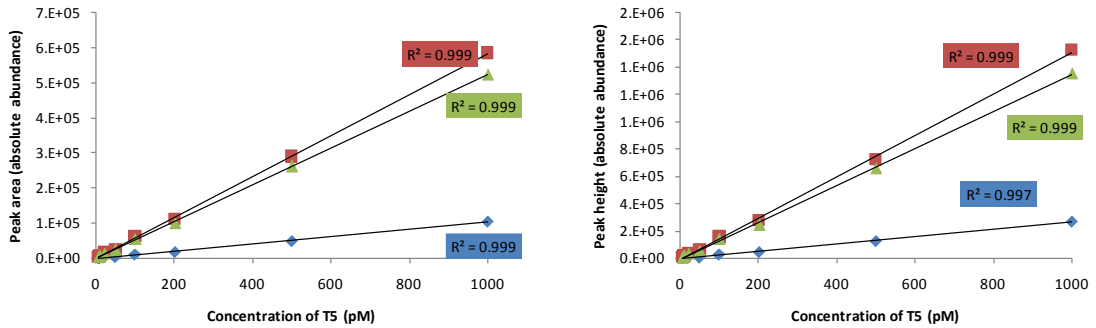
Concentration of T5 (pM)	Peak area (conventional-flow LC-MS)			Peak height (conventional flow LC-MS)		
	m/z 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363	*m/z* 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363
100	36	253	150	124	703	525
200	88	512	391	305	1771	1315
500	190	1108	911	641	3518	3100
1000	415	1866	1631	1406	6262	5686

Figure 3.4-12: Linearity of synthesised P-III-NP T5 peptide standards analysed by conventional-flow LC-MS using absolute peak area and height.



Concentration of T5 (pM)	Peak area (micro-flow LC-MS)			Peak height (micro-flow LC-MS)		
	m/z 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363	*m/z* 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363
20	86	688	394	249	2007	1384
50	128	1142	818	509	3703	2776
100	254	1729	1367	952	6154	5059
200	569	3989	3580	2030	11838	9605
500	1636	10765	8426	5883	37560	30148
1000	2844	20328	15982	10500	67368	54921

Figure 3.4-13: Linearity of synthesised P-III-NP T5 peptide standards analysed by micro-flow LC-MS using absolute peak area and height.



Concentration of T5 (pM)	Peak area (nano-flow LC-MS)			Peak height (nano-flow LC-MS)		
	m/z 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363	*m/z* 534 -> 628	*m/z* 534 -> 448	*m/z* 534 -> 363
5	539	2948	2690	1307	6900	6165
10	1093	6126	5540	2815	15524	13601
20	2792	15527	13819	6794	38912	34251
50	4604	25541	22746	4604	63863	56810
100	11436	63814	56589	28735	158157	142217
200	18671	111934	100152	46345	276534	246594
500	49826	289830	262599	125041	726076	662545
1000	103726	584880	524313	270923	1525148	1358857

Figure 3.4-14: Linearity of synthesised P-III-NP T5 peptide standards analysed by nano-flow LC-MS using absolute peak area and height.

3.4.3 *Effect of matrix on P-III-NP LC-MS analysis*

Despite the sensitivity of the conventional-flow LC-MS method, P-III-NP peptides originating from trypsin digested bovine P-III-NP (~100 µg/mL, 2.5 µM) in HSA (50 mg/mL, 757 µM) was not observed using conventional LC-MS methods. For the *b*T1 peptide, limited DTT and IDA may affect observance of peptide, however these reagents do not affect the yield of T5. A high concentration of trypsin (10 µg) was used to facilitate the digestion of P-III-NP and HSA, which were both present at relatively high concentrations. Thus, based on this observation, serum P-III-NP will need to be separated from albumin (and possibly other serum proteins) prior to trypsin digestion to enable LC-MS analysis.

In serum, trypsin is not selective to digesting P-III-NP, hence, the LC-MS tolerance of P-III-NP peptides for HSA (the most abundant serum protein, basal conc. 35-50 mg/mL) tryptic peptides was investigated. At the HSA peptide concentrations (0.1-100 µg) injected, P-III-NP peptides (100 pM) were not observed using conventional-flow LC-MS methods. For the samples analysed by nano-flow LC-MS, peptides (100 pM) were observed in 0.1 -1 µg HSA (2-20 µg/mL, 30-300 nM) peptide matrix (see Figure 3-17). Increasing the matrix concentration ≥ 5 µg/50 mL (100 µg/mL, 1.5 µM) resulted in complete suppression of P-III-NP peptides in the MS. Hence, to analyse T5 and T1peptides originating from serum P-III-NP by the developed LC-MS methods, an HSA depletion efficiency > 99.6 % is necessary.

3.4.4 *Activity of immunoassay for P-III-NP peptides and protein*

Using the Orion P-III-NP assay, activity for intact bovine P-III-NP sample (~50 ng/mL) was 3 ng/mL against the standard curve. This is low, albeit not unexpected, as the P-III-NP in the original sample may have degraded due to age and storage conditions. For the monomeric single stranded protein and tryptic peptides (*h*T1, *b*T1 and T5) no activity was observed. These results would indicate that the P-III-NP antibody contained within the kit is specific for the trimeric intact protein. Hence, this assay

cannot be used as an independent quality control check for P-III-NP concentration throughout sample preparation for LC-MS.

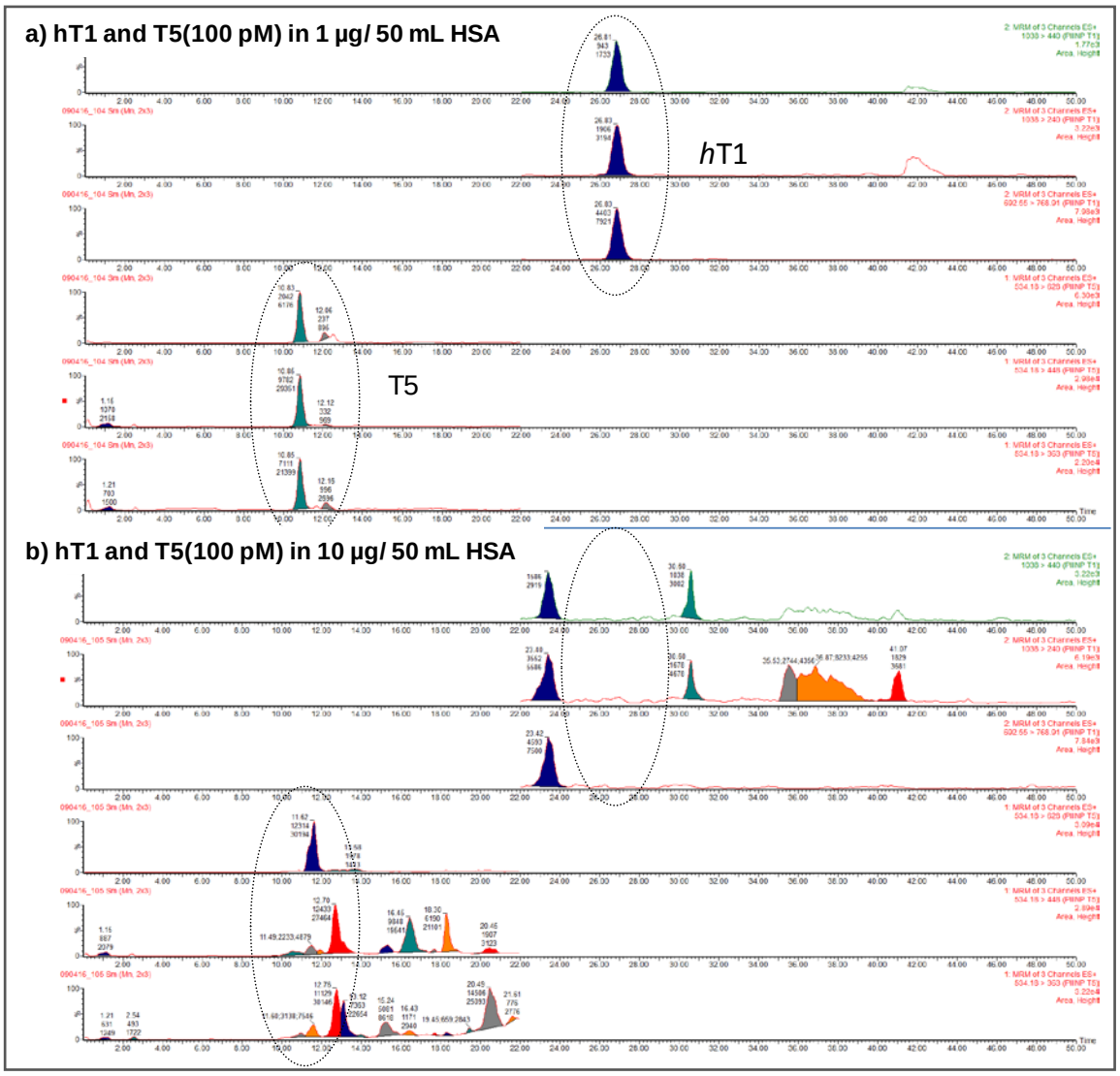


Figure 3.4-15:Extracted ion chromatograms of P-III-NP peptides (*h*T1 and T5, 100 pM) in HSA trypsin peptides in a) 1 µg/50 µL (300 nM) and b) 10µg/ 50 µL (3 µM) analysed by nano-flow LC-MS. The elution region of *h*T1 and T5 peptides are highlighted.

3.5 **Summary**

Sensitive methods capable of analysing P-III-NP at basal serum levels (1- 5 ng/mL, 25-125 pM) have been developed using conventional-, micro- and nano-flow LC-MS set-ups on a highly sensitive triple quadrupole MS. However the presence of albumin (and other serum proteins) limit the capabilities of these methods. At concentrations > 20 µg/mL (300 nM) HSA, tryptic peptides completely suppress the signal of P-III-NP

hT1 and T5 peptides (100 pM) in the MS. To overcome this P-III-NP needs to be separated from other serum proteins (especially albumin) prior to trypsin digestion, so as to enable LC-MS analysis. The antibody specificity of the Orion Diagnostica P-III-NP kit for intact P-III-NP limits the use of the assay as an independent concentration check for P-III-NP during sample preparation for LC-MS. In Chapter 4 methods to separate P-III-NP from HSA are investigated, so as to develop an appropriate sample preparation protocol for the analysis of P-III-NP from human serum.

Chapter 4 *: Sample preparation and effect of carrier proteins*

4.1 **Overview**

In the previous chapter we highlighted the need to separate HSA (and other serum proteins) from P-III-NP in blood prior to trypsin digestion so as to analyse by LC-MS. Similarity in size, but vast difference in concentration of P-III-NP (42 kDa, 1-5 ng/mL) and HSA (66 kDa, 35-50 mg/mL) make separation difficult. Here the potential of common proteomic separation techniques (i.e. protein precipitation, molecular weight cut-off filtration, and immunocapture) are assessed for separating P-III-NP from serum proteins. The efficiency of HSA removal and P-III-NP recovery of these approaches have also been determined.

4.2 **Introduction**

Circulating intact P-III-NP related to GH administration is obtained from blood. The incomplete definition of the blood proteome results from its abundance in albumin (55 %), wide dynamic range in protein concentrations (in the region of 10-12 orders, g/L- pg/L) and the presence of glycoproteins [159]. The complexity and dynamic range of proteins presented in this matrix complicates MS detection; and current MS instruments cannot analyse over this range without pre-fractionation of the sample [160]. As a result of the presence of albumin (see Figure 4-1), significant MS interference has been observed in the analyses of low abundant proteins like P-III-NP from serum [161-164]. Hence, for our purpose, the separation of HSA from P-III-NP is the primary aim of sample preparation. To achieve this and to simplify the serum proteome either HSA depletion or P-III-NP enrichment is necessary.

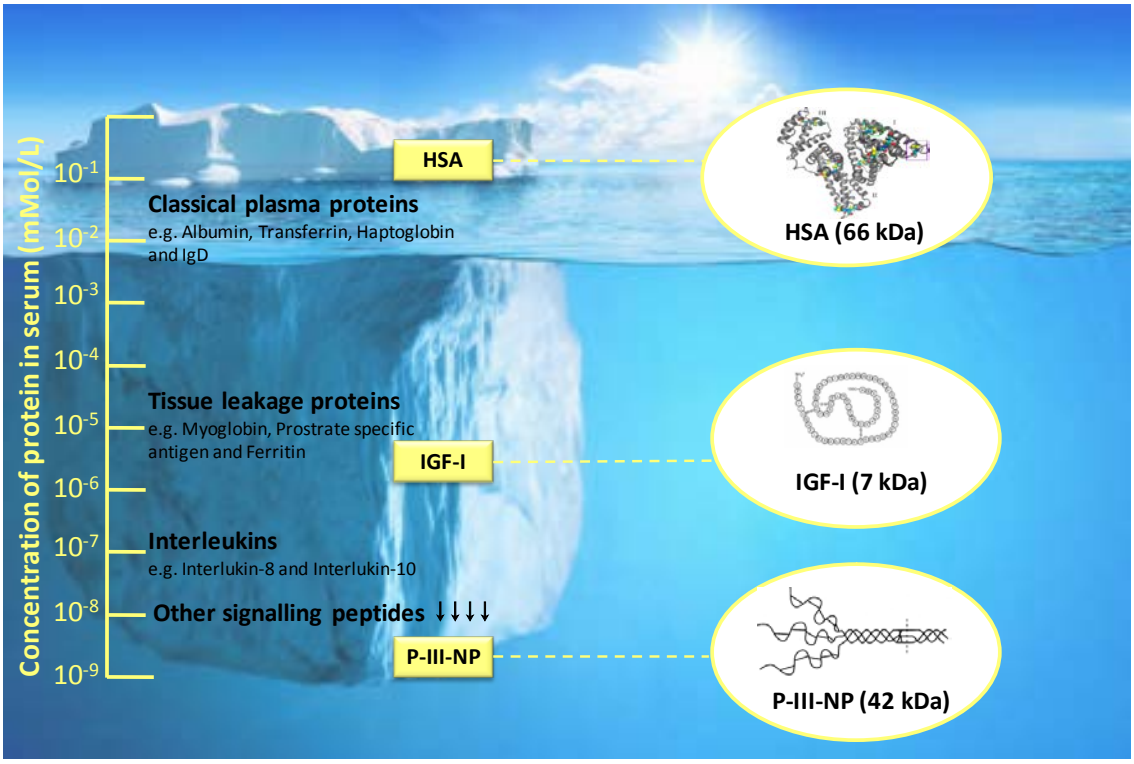


Figure 4.2-1: Illustrative iceberg model to show the dynamic range of plasma proteins in serum where the concentrations of the classes of proteins shown are summarised from [159]. HSA and GH biomaker proteins, IGF-I and P-III-NP, are highlighted to demonstrate their basal concentration range against the human blood proteome.

4.2.1 *Disulphide bonds*

In Chapter 2 some of the structural properties of P-III-NP are discussed. Here the tropocollagen or "collagen-like" nature of the protein is further discussed, with specific focus on the disulphide bonds, to investigate whether this function can be used to achieve a mass separation from albumin that would potentially enable the physical separation of the two proteins in serum. Disulphide covalent bonds are chemical bonds that occur between two sulfhydryl or thiol groups. In proteins these bonds occur between cysteine residues, which have a thiol group on its side chain that is easily oxidised to form a sulfhydryl (-S-S-) bridge, shown in Figure 4-2 below. The formation of disulphide bonds between the correct pairs of cysteine residues is essential for the folding and stability of many proteins [165].

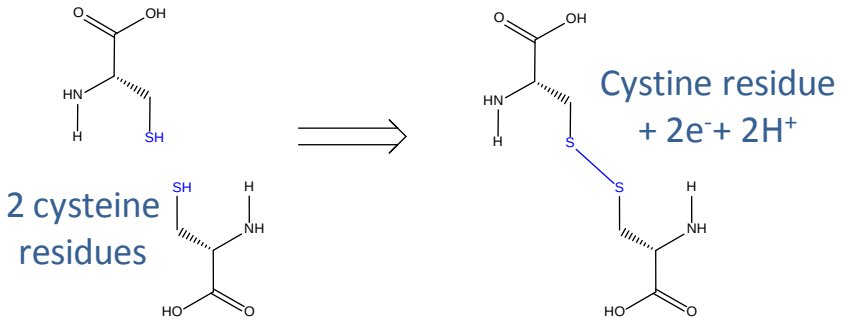


Figure 4.2-2: Disulphide bond formation due to the oxidation of two cysteine residues to for a cystine residue.

4.2.1.1 P-III-NP disulphide bonds

In the Col 1 domain of P-III-NP, five intra-chain disulphide bridges on each of the pro α1-chains give a globular structure to this region making it hydrophilic [129]. In this conformation, the hydrophobic amino acid residues are folded into the protein exposing the more polar residues outward where they can have dipole-dipole solvent interactions, thus attributing solubility to the protein. Whilst in the Col 2 domain six inter-chain disulphide bonds between the three pro α1-chains exist. These bonds act as a nucleus for the left-handed twisting of the collagen-like region in the Col 3 domain [166] and stabilises the protein structure.

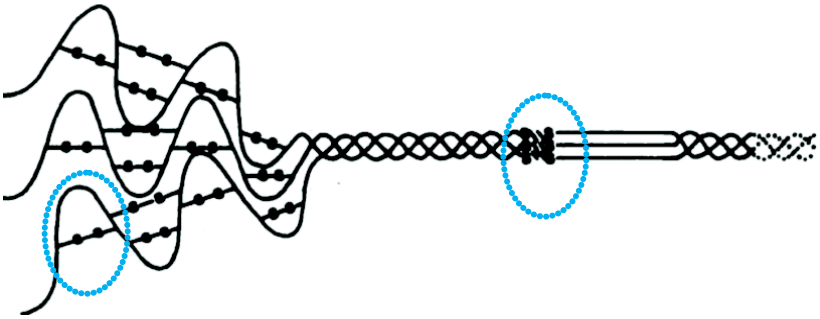


Figure 4.2-3: Schematic illustration of the trimeric P-III-NP protein. Disulphide bonds that connect the three identical pro α-1 chains in the Col1 and Col2 domain are shown using a ball and stick, two are highlighted with blue circles. (Image adapted with permission from [70]).

These disulphide bonds can be reduced in P-III-NP with the application of heat (> 55 °C) which denatures the protein; however, upon cooling the protein to approximately 20 °C, complete renaturation is seen to occur if these inter-chain disulphide bonds are not subsequently blocked [66]. With reduction and blocking of the disulphide bridges (reaction shown in Figure 4-4) within P-III-NP, the tertiary structure

of the protein is permanently destroyed to generate three identical partial pro-collagen α1 (III) chains, each consisting of 130 aa with a ~14 kDa MW.

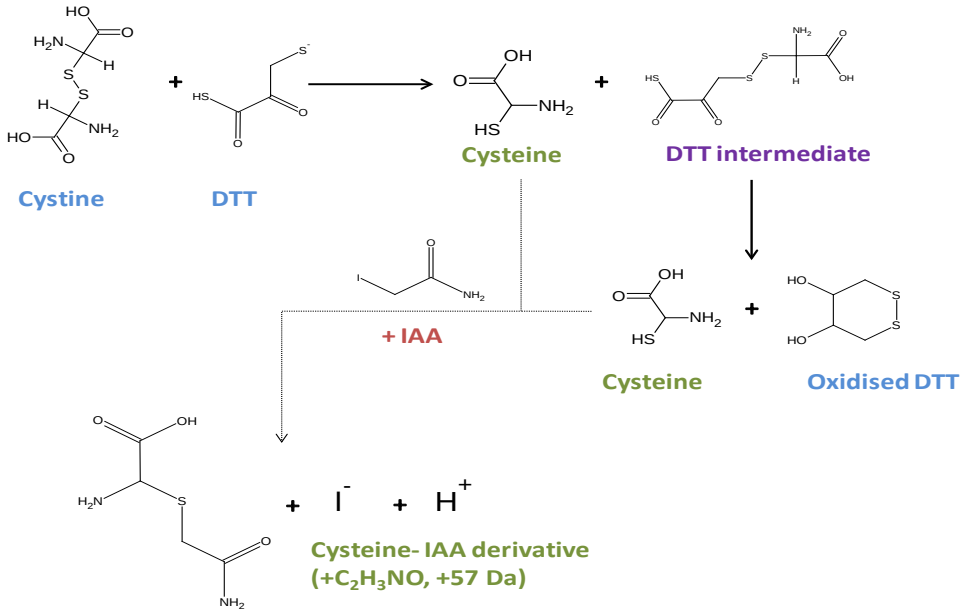


Figure 4.2-4: Reduction of a cystine disulphide bond with dithiothreitol (DTT) following the blocking of sulfhydryl group by alkylating with iodoacetamide (IDA) to form a carbamidomethyl cysteine IDA derivative.

4.2.1.2 Albumin disulphide bonds

HSA is a 585 aa protein with an α-helix heart structure divided into three homologous domains, shown in Figure 4-5 below. Containing thirty-five cysteine residues, seventeen disulphide bonds exist in HSA to leave a free C residue. This free thiol (C-34) can potentially form inter-molecular disulphide bridges with other HSA molecules or thiol-containing substances in plasma. However, due to steric hindrance affecting the accessibility of C-34, the majority of serum HSA *in vivo* remains as HSA-SH [167].

Of the 17 disulphide bonds in HSA formed by the other 34 cysteine residues, 16 are "double S-S bonds" in that they exist between residues located beside each other within the molecule. Reduction and alkylation of HSA results in the unfolding of the protein's tertiary structure to leave most of the secondary structure intact except for a few α-helices [168]. No hydrolysis of amide bonds occurs in this process and so

considering that all disulphide bonds are intra-molecular, the MW of HSA remains > 66 kDa after reduction.

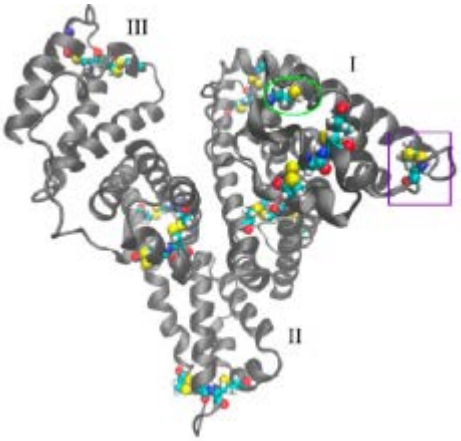


Figure 4.2-5: Schematic of HSA showing its three homologous domains (I, II and III). Disuphide bonds (e.g. highlighted with purple box)between cysteine residues (yellow). A single cysteine group (C-34) that remains free or does not form a disuphide bond is highlighted by the green circle. (Figure reprinted with permission from [168]).

4.2.2 *Depletion methods*

High abundant proteins (e.g. albumin and IgG) account for 90 % of the total protein count in serum [159]. With the achievable differences obtained in the MW and isoelectric point (pI) of HSA and P-III-NP (after disruption of their tertiary structure) investigation into the application of classical proteomic depletion approaches to reduce the complexity of the serum are considered. These include organic solvent protein precipitation (ppt) [57, 169], ultrafiltration [91, 162], albumin immunodepletion (or affinity capture) [161, 170, 171].

4.2.2.1 Protein precipitation

Precipitation of proteins in solution results from the change in solubility due to a change in the chemical environment that decreases the solvation potential. Polar interactions with solvent, ionic interactions with salts and the repulsion of like electrostatic forces within the protein leads to solubility of the protein [90]. Hence, at the isoelectric point (pI) of a protein where there is no net charge, proteins are insoluble in solution and will precipitate. At a pH above a protein's pI there is a net negative charge; below the pI a net positive charge exists.

The addition of organic solvent (e.g. acetonitrile, acetone or ethanol) to serum precipitates HMW (high molecular weight) proteins to leave the more soluble low MW (LMW) fraction in solution. Addition of organic solvents displaces the water molecules around the hydrophobic regions of the proteins and lowers the dielectric constant of the solution resulting in an increase in the attraction of charged molecules which promotes electrostatic interactions within the protein to result in precipitation. Acetonitrile (ACN) has been found to be the most reproducible and effective at removing HSA [90]. During precipitation the tertiary structure of a proteins is removed , as it is in a denatured state, hence association with other proteins (e.g. binding proteins) will be lost. At concentrations > 20 %, ACN has been shown to disrupt the association of proteins to HSA [160]. Hence, as a relatively cheap approach to fractionation it is not surprising that ACN depletion is often used in proteomics [57, 169].

Alternatively salts (e.g. ammonium sulfate), metal ions (e.g. zinc sulfate), or acids (e.g. trichloroacetic acid) can be used to precipitate proteins. Like organic solvents the addition of salts to protein solutions removes water molecules from the protein surface by preferential binding to result in ppt as a result of aggregation due to hydrophobic interactions. Metal ions compete with solution protons for coordination of binding sites on the protein to effectively lower both the solution pH and protein's pI to result in ppt. Acids form insoluble salts with positively charged aa groups within the protein, thus increasing interaction with solvent to increase insolubility and result in ppt. Although capable of ppt these approaches are less frequently used compared with organic solvents, due to their lack of ease in compatibility with LC-MS where they affect ionisation efficiencies.

4.2.2.2 Ultrafiltration

A gentler approach to protein fractionation is ultrafiltration, which uses a size exclusion mechanism to separate LMW proteins from HMW proteins with a molecular sieve. Some users claim that separation of proteins with ultrafiltration is only observed when the MW difference to be separated is > 10 fold, however this is rarely the case [172].

Primarily, separation is achieved by difference in MW, however, solute-solute interactions, solvent-membrane interactions, mode of solute transport (connective or diffusive) and concentration also need to be considered.

Serum samples are often denatured to eliminate solute-solute interactions especially for albumin and the LMW proteins it potentially transports [160]. Membrane-solute interactions can be minimised by either passivating the membrane to eliminate exposed non-specific binding sites, or by changing the solution composition with the addition of proteins (e.g. albumin), salts (e.g. sodium dodecyl sulfate) or surfactants (e.g. Tween-20) [173]. Clogged membranes can result from loading too concentrated samples like serum onto the filter; hence, samples are to be diluted. Slower speed and longer centrifugal spin times (than those recommended by manufacturers) can potentially increase the recovery of LMW proteins, however, the passage of HMW proteins through the membrane also increases [174].

4.2.2.3 HSA depletion columns

High selectivity of HSA depletion using affinity capture columns has resulted in the common use of depletion columns in the sample preparation of serum for the analysis of low abundance proteins (LAPs) [161, 171]. Several mechanisms for binding exist, these include Cibacron Blue F3GA [175] or specific antibodies [176].

It is not surprising that the use of monoclonal antibodies to capture HSA from serum is highly specific. However, the efficiency of this approach is limited in the volume of sample that can be applied to this system as a result of the low binding capacity of antibodies [161]. To overcome this, samples can be split into several aliquots for HSA depletion and pooled after separation, however the expense associated with these systems (especially in culturing the antibodies) does not make this feasible. Cibacron blue on the other hand is a cheaper approach to albumin depletion in that it mimics nicotinamide adenine and purine dinucleotides by binding proteins containing a dinucleotide fold with high affinity interaction [177]. Lollo *et al.* [170] state that non-selective binding of proteins can occur through a complex combination of electrostatic,

hydrophobic and hydrogen bonding interactions with the planar ring structure of Cibacron Blue or its negatively charged sulfate groups. Despite this approach, having a relatively high load capacity (compared with antibody based systems), lack of selectivity of its secondary binding results in the unpredictable loss of proteins which can eliminate its advantage of use for LAPs in serum.

4.2.3 ***Immunocapture***

Contrary to albumin depletion, the enrichment of target proteins (or corresponding proteolytic peptide) is an alternative approach to simplify serum samples to enable the analysis of LAPs. Using an antibody-based sample clean-up approach like immunocapture or immunoaffinity enrichment has enabled the sensitive analysis of serum LAPs by LC-MS [93, 178]. This is mainly due to the affinity of the antibody used to target the desired protein during sample clean-up. Although some degree of specificity for the antibody is required, unlike immunoassays, polyclonal antibodies are appropriate for immunocapture coupled to bottom-up LC-MS, since the specificity of LC-MS analysis is dependent on the appropriate selection of proteolytic peptides with unique SRM transitions. Antibody targets can be proteins or peptides, although proteins are preferred as they have stronger affinity [94]. The availability and cost to culture such antibodies is the main limitation to this approach, thus matrix clean-up in proteomics is still heavily dependent on depletion approaches.

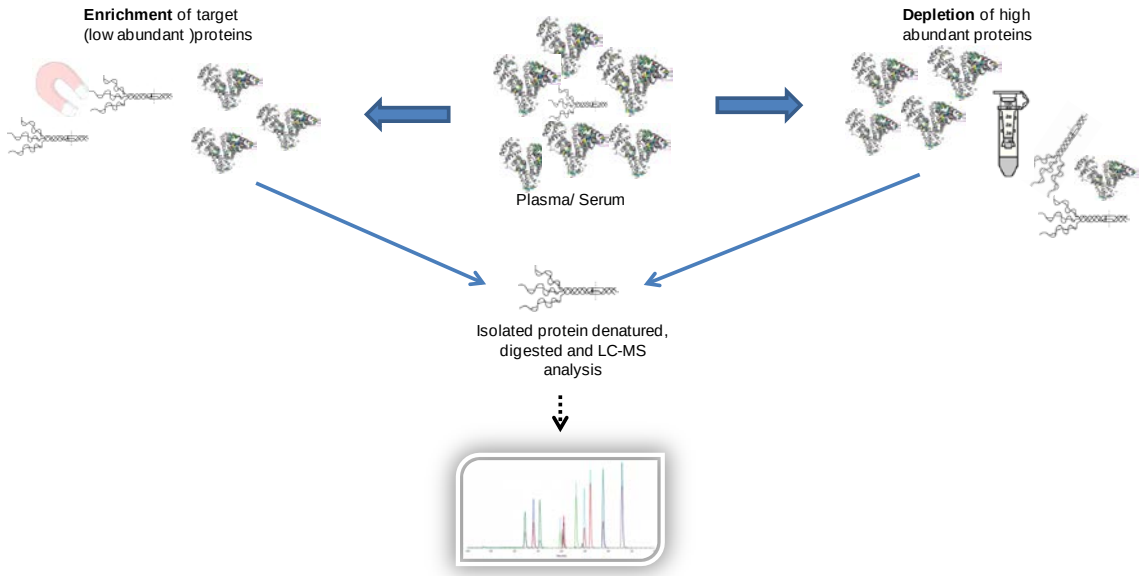


Figure 4.2-6: Schematic to show the possible sample preparation workflows that should simplify the serum proteome to enable analysis of low abundant proteins such as P-III-NP.

4.3 **Materials and Methods**

The source of materials where unmentioned can be found in section 2.3 and section 3.3. Acetic acid glacial (Analytical Reagent grade) and ammonium acetate (Certified AR for Analysis) were purchased from Fisher Scientific (Loughborough, UK). Trifluoroacetic acid (99 %, Reagent Plus®) and Tween-20 were purchased from Sigma Aldrich (St. Louis, MO, USA).

4.3.1 ***Separation of HSA and P-III-NP by protein precipitation***

Acetic acid (1 %, 200 µL) was added to bovine P-III-NP (10 µg/mL in 50 mg/mL HSA, 100 µL) and HSA (50 mg/mL, 100 µL) samples and vortexed (10 sec). To each sample acetonitrile (400 µL) was added before gently mixing for 10 min at 25 °C at 900 rpm. Samples were then centrifuged for 7 min at 13,000 g, before transferring the supernatant to a clean Lobind® tube. The supernatants were evaporated at 45 °C for approximately 2 h. Samples were reconstituted in ammonium bicarbonate (100 mM, 50 µL) before digesting with trypsin (2 µg) using the protocol described in section 2.3.4. Samples were analysed by conventional-flow LC-MS using the LC parameters described in section 3.3.2 whilst monitoring the SRM transitions listed in Table 4-1 below.

Table 4-1: Collision energies for SRM (selected reaction monitoring) transitions monitored for *b*P-III-NP, developed in the absence of synthesized *b*T1 peptide developed using a Waters Xevo® TQ-S MS.

P-III-NP Peptide	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (arbitrary units)
***b*T1**	688	539	13
	688	704	10
	688	761	10
T5	534	363	20
	534	448	13
	534	628	27

4.3.1.1 Monomeric single stranded P-III-NP

To convert intact bovine P-III-NP (10 µg/mL in 0.1 % HSA, 100 µL) to its monomeric single-stranded form, disulphide bonds were removed by reduction (DTT) and alkylation (IDA) using the protocol described in section 2.3.2.1. Once derivatised, the

monomeric P-III-NP was added to carbamidomethylated HSA (50 mg/mL, 100 µL). The sample was then precipitated using the protocol described above, before analysing by conventional-flow LC-MS.

To determine whether losses of P-III-NP could be due to the protein being trapped by the high concentration of HSA normally found in matrix, the experiment was repeated with a smaller volume of HSA (50 mg/mL, 10 µL). Separation of P-III-NP peptides and serum peptides were increased by decreasing the gradient slope of the LC method using the conditions described in Table 4-2 below.

Table 4-2: Extended LC gradient conditions used to achieve additional separation of P-III-NP peptides from serum peptides using conventional-flow LC.

Time (min)	Flow (mL/min)	% A	% B
0.00	0.3	95	5
3.00	0.3	93	7
23.00	0.3	92	8
24.00	0.3	89	11
44.00	0.3	88	12
45.00	0.3	10	90
45.01	0.3	95	5
47.00	0.3	95	5

4.3.1.2 Monomeric single stranded P-III-NP at pH 5.97

To a sample of monomeric bovine P-III-NP (10 µg/mL in 50 mg/mL carbamidomethylated HSA, 10 µL), ammonium acetate (100 mM, 90 µL) was added. The pH of the sample was adjusted to 5.97 with acetic acid (1 %), whilst measuring with a Benchtop pH meter (Orion™ 3star, Thermo Scientific), before adding acetonitrile (300 µL) and vortexing for 5 sec and incubating at 25 °C for 10 min to facilitate protein precipitation. The supernatant was removed, to the remaining pellet acetonitrile (300 µL) was added before vortexing (5 sec) and centrifuging at 4 °C for 10 min at 16,100 g. Supernatants were pooled and evaporated for approximately 3 h at 45 °C in a Speedvac, before reconstituting in ammonium bicarbonate (100 mM, 50 µL) and digesting with trypsin (2 µg) using the protocol described in section 2.3.4. Samples were analysed by conventional-flow LC-MS, using the extended gradient LC conditions.

4.3.2 *Separation of HSA and P-III-NP using molecular weight cut-off filtration*

To reduce losses resulting from non-specific binding to molecular weight cut-off (MWCO) centrifugal filters (Merck Millipore; Cork, Ireland), columns were passivated with Tween-20 as follows. Tween-20 (5 %, 500 µL) was loaded into each MWCO filter and incubated for 1 h at room temperature (approx. 25 °C) before removing with a pipette. The filter was washed with sterile distilled water (500 µL) to remove residual Tween-20 and centrifuged for 10 min at 14,000 g. This washing step was repeated twice more. Once passivated, MWCO filters were either used immediately or stored at 4 °C. For storage, sterile distilled water (100 µL) was loaded into the filter to retard bacterial growth.

Intact bovine P-III-NP (10 µg/mL in 50 mg/mL HSA, 100 µL) and HSA (50 mg/mL, 100 µL) were reduced (DTT) and alkylated (IDA) using the protocol described in section 2.3.2.1. To each sample, ammonium bicarbonate (100 mM, 200 µL) was added before loading into a passivated 30 kDa MWCO centrifugal filter. Samples were centrifuged for 10 min at 14,000 g, before collecting the filtrate and evaporating to dryness at 45 °C for ~4 h. Samples were then reconstituted in ammonium bicarbonate (100 mM, 50 µL) and digested with trypsin (2 µg) at 37 °C overnight whilst shaking at 1,400 rpm. To each sample, formic acid (2 %, 10 µL) was added before transferring to an autosampler vial and for analysis by conventional-flow LC-MS, using the extended LC gradient conditions shown in Table 4-2.

4.3.2.1 Efficiency of Albumin depletion

Carbamidomethylated HSA samples (50 mg/mL, 100 µL) were albumin depleted using protein precipitation and MWCO (10 and 30 kDa) filtration protocols (sections 4.3.1 and 4.3.2). Depleted samples and a HSA standard (500 µg/mL, 100 µL) were digested with trypsin (2 µg) overnight at 37 °C with shaking at 1,400 rpm. To stop trypsin activity, formic acid (2 %, 10 µL) was added to each sample before evaporating to dryness at 45 °C for approx. 6 h. Samples were then reconstituted in formic acid (0.3 %) and

acetonitrile (5 %) solution (50 µL) containing Substance P (37 nM). Samples were analysed by conventional-flow LC-MS using targeted SRM methods for HSA tryptic peptide, VFDEFKPLEEPQNLIK (with Substance P as reference) to assess the efficiency of HSA depletion.

Analysis was done on a Waters Acquity® UPLC® coupled to a Waters Xevo® TQ-S MS interfaced with a conventional ESI ion source using the MS tune parameters listed in section 3.3.1. Gradient LC separation of digested HSA (10 µL) was achieved on an Acquity® UPLC® BEH 130Å (C18, 1.7 µm, 2.1 x 100 mm) column with mobile phase A, 0.3 % formic acid in water, and B, 0.3 % formic acid in acetonitrile and a flow rate of 300 µL/min. Mobile phase B was held at 20 % for 3 min before increasing to 50 % in 7 min, then to 95 % in 2 min and held for a further minute. Mobile Phase B was then immediately returned to 20 % and held for 2 min to re-equilibrate the column. The SRM transitions monitored for the selected HSA peptide and substance P are shown in Table 4-3 below.

Table 4-3: Collision energies for SRM transitions acquired for trypsin generated HSA peptide, VFDEFKPLEEPQNLIK, and substance P for conventional-flow LC-MS methods developed using a Waters Xevo® TQ-S MS.

Peptide	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (arbitrary units)
HSA peptide VFDEFKPLEEPQNLIK	682	219	25
	682	247	17
	682	900	20
Substance P	449	253	15
	449	543	15
	449	600	10

4.3.2.2 Depletion of albumin from pooled serum samples

Albumin was depleted from serum samples (100 µL) in triplicate using protein precipitation, MWCO filtration (30 kDa), combined protein precipitation then MWCO (30 kDa) filtration and double MWCO (30 kDa) procedures. Once depleted, samples were digested with trypsin (2 µg/mL) and evaporated using the conditions described above. One sample from each triplicate was reconstituted in formic acid (0.3 %) and acetonitrile (5 %) solution (50 µL) containing Substance P (37 nM), and analysed for the targeted HSA peptide by conventional-flow LC-MS. The second sample was

reconstituted in formic acid (0.3 %) and acetonitrile (5 %) solution (50 µL) containing 100 pM *h*T1 and T5 peptides, and analysed for the targeted P-III-NP peptides by nano-flow LC-MS using the conditions described in sections 3.3.1 and 3.3.2.3. The remaining sample was reconstituted in formic acid (0.3 %) and acetonitrile (5 %) solution (50 µL) and analysed using the data-dependent method detailed in section 2.3.3.2.3 to identify proteins (with a threshold > 80 %) remaining in depleted matrix.

4.3.3 ***Recovery of P-III-NP from double MWCO filtration***

Bovine P-III-NP (~100 ng/mL, 100 µL) in HSA (50 mg/mL) and without HSA, and HSA (50 mg/mL, 100 µL) samples were reduced and alkylated using the protocol described in section 2.3.2.1. The samples containing HSA were then filtered with a 30 kDa MWCO filter twice, using the protocol described in section 4.3.2 above. All samples were then evaporated to dryness at 45 °C for approximately 6h, before reconstituting in ammonium bicarbonate (100 mM). Samples were then digested overnight with trypsin (2 µg) at 37 °C whilst shaking at 1,400 rpm. To stop trypsin activity formic acid (2 %, 10 µL) was added before transferring samples to an autosampler vial and analysed by nano-flow LC-MS using the conditions described in sections 3.3.1 and 3.3.2.3 to assess the recovery of P-III-NP from MWCO filtration.

4.3.3.1 Effect of potential carrier proteins on P-III-NP analysis

To improve recovery of P-III-NP from double MWCO filtration, the ability of abundant proteins tetracosatide (ACTH 1-24; Sigma Aldrich, St. Louis USA), Substance P, and recombinant human insulin (Sigma Aldrich, Buchs, Switzerland) to act as a "carrier" for *h*T1 and T5 during sample preparation, was investigated.

Tetrocosactide, substance P and insulin samples (each at 150 ng/mL, 100 µL) were digested with trypsin using the protocol described in section 2.3.4, except at reconstitution stage formic acid (0.3 %) and acetonitrile (5 %) solution (50 µL) containing 100 pM *h*T1 and T5 peptides was added to the tubes. Samples were analysed by nano-flow LC-MS using the conditions described in sections 3.3.1 and

3.3.2.3. The action of insulin as a carrier for P-III-NP was further investigated using the double MWCO filtration.

To a carbamidomethylated sample of *b*P-III-NP (100 ng/mL in 50 mg/mL HSA) insulin (150 ng/mL, 200 µL) was added before transferring to a MWCO (30 kDa) filter for separation. Once double MWCO filtered and digested with trypsin (2 µg), samples were analysed by nano- flow LC-MS. This experiment was repeated, however, insulin (150 ng/mL) was used to passivate MWCO filters instead of Tween-20. Samples were analysed by nano-flow LC-MS using the conditions described in sections 3.3.1 and 3.3.2.3 with modified *b*T1 SRM transition listed in Table 4-4 below. To assess whether insulin improves the recovery of P-III-NP (compared to digested *b*P-III-NP (100 ng/mL in 0.1% HSA) standard) *b*T1 and T5 signals were compared to those obtained without a "carrier".

Table 4-4: Initial collision energies for SRM (selected reaction monitoring) transitions monitored for *b*P-III-NP; developed with synthesized *b*T1 peptide developed using a Waters Xevo® TQ-S MS.

P-III-NP Peptide	Precursor ion (*m/z*)	Product ion (*m/z*)	Collision energy (arbitrary units)
***b*T1**	688	761	15
	688	440	20
	688	369	20
	688	240	20

4.3.4 ***Separation of HSA and P-III-NP using immunocapture***

For the following protocol both anti-P-III-NP monoclonal antibodies and Cal 0 (buffered solution containing bovine proteins, preservatives and yellow-orange dyes) were sourced from a P3NP-EL-US immunoassay kit (CisBio Bioassays, Codolet, France).

Bovine P-III-NP (100 ng/mL in 50 mg/mL HSA), HSA (50 mg/mL) and Cal 0 were brought to room temperature (18-25 °C). In duplicate, each sample (100 µL) was pipetted into appropriate microtitre wells containing immobilised anti-P-III-NP mouse monoclonal antibody. A second mouse monoclonal anti-P-III-NP antibody (linked to horseradish peroxidase, mouse immunoglobulins, stabilisers and preservative, 100 µL) was added to each well before covering with adhesive film and incubating overnight at

4 °C whilst shaking at 700 rpm. The adhesive film was removed and the contents of the wells were dispelled before washing with PBS buffer containing 0.3 % Tween-20 (300 µL) three times. To each well elution solution (200 µL), containing tetracosactide (ACTH 1-24, 15 mg/mL), acetonitrile (33 %) and trifluoroacetic acid (0.4 %) was added before incubating at 60 °C for 5 min whilst shaking at 900 rpm. Samples were then transferred into labelled LoBind Eppendorf® tubes and digested with trypsin (2 µg) as described in section 2.3.4. Samples were analysed with nano- and micro-flow LC-MS using the conditions described in sections 3.3.1 and 3.3.2.

4.4 **Results and Discussion**

Using the developed conventional LC method described in section 3.3.2.1, a potential interfering early eluting peak at ~t_R 4.5 min (Figure 4-7) can be seen close to the expected region for the P-III-NP T5 peptide (t_R 5 min). In the data dependent results for trypsin digested HSA, numerous peptides were observed in the eluting region of T5 hence to better separate and eliminate potential ion suppression of our P-III-NP peptides the slope of the LC gradient in this region was decreased in future experiments. At the time of analysis, a synthesised standard for the *b*T1 peptide was not available, hence the SRM transitions used were inferred from the fragmentation pattern of *h*T1 with the incorporation of aa sequence differences. A tryptic digest of *b*P-III-NP analysed by LC-MS using these transitions gave two characteristic peaks for *b*T1 (approximately 1 min apart), this doublet peak was used to identify the *b*T1 peptide.

4.4.1 ***Separation of HSA and P-III-NP by protein precipitation***

Separation of P-III-NP and HSA was not achieved by protein precipitation. In all the ppt experiments P-III-NP was observed in the precipitation fraction. This observation is most likely due to insolubility of the protein in acetonitrile, however it could also be due to the very abundant HSA, which is expected to precipitate, trapping the P-III-NP in its clusters during the precipitation process.

In attempted separation of P-III-NP (10 µg/mL) from HSA (50 mg/mL) a small amount of P-III-NP was observed in the acetonitrile supernatant (see Figure 4-7), however, most of the protein is in the precipitate along with HSA. The efficiency of ppt is not expected to be 100 % [90], hence observance of P-III-NP in the supernatant is likely to be a result of the incomplete precipitation of the protein. With a reduction in the concentration of P-III-NP and HSA (diluted 1 in 10) no P-III-NP was observed in the supernatant fraction after ppt. The precipitate contained a small amount of P-III-NP T5 peptide. No *b*T1 peptide was observed in either fraction. Absence of P-III-NP in the supernatant of the diluted sample (Figure 4-8) suggests that P-III-NP is in fact

precipitated by acetonitrile and not trapped by HSA; the low concentration of the recovered analyte in the precipitate is probably due to loss during sample handling, or ion suppression by HSA peptides.

The effect of the size and tertiary structure of P-III-NP on precipitation was investigated with the attempted ppt separation of trimeric and monomeric *b*P-III-NP (both 10 µg/mL) from HSA (50 mg/mL). Unfortunately, both P-III-NP variants were insoluble in acetonitrile and precipitated, clearly demonstrating that these variables do not have an effect on ppt.

Based on amino acid composition, the pI was calculated for P-III-NP (3.97) and HSA (5.67). Where the pH of a solution equates to a protein's pI, proteins are less soluble due to there being a net zero charge on the functional groups, which should readily facilitate precipitation. With this in mind, selective precipitation of HSA was attempted at pH 5.97. At this pH albumin should remain uncharged and thus insoluble in the aqueous solution (supernatant), whilst P-III-NP should be charged and interact with the solvent. Unfortunately, the proteins co-precipitated, thus separation was not achieved. The unexpected precipitation of P-III-NP at pH 5.97, which should have a net negative charge and interact with the solvent, could be due to the incorrect assignment of pI as a result of the incomplete characterisation of the protein as the presence of PTMs will also affect the solubility of the protein.

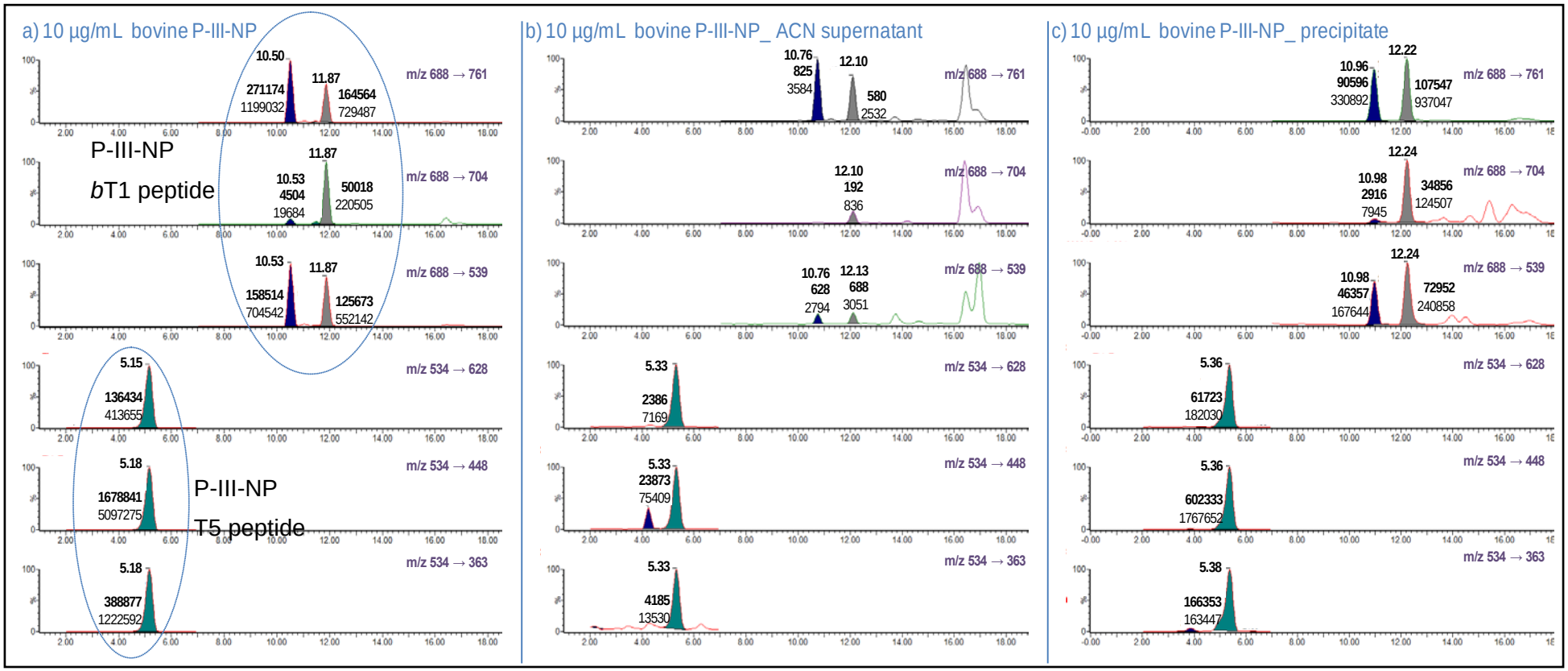


Figure 4.4-1: Chromatograms of a *b*P-III-NP sample (10 µg/mL) without HSA (a), and the acetonitrile and precipitate fractions collected from an ACN ppt of *b*P-III-NP (10 µg/mL) in 50 mg/mL human serum albumin (b and c, respectively), analysed by conventional-flow LC-MS after digesting overnight with trypsin. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *b*T1 and T5 peptides have been highlighted in a).

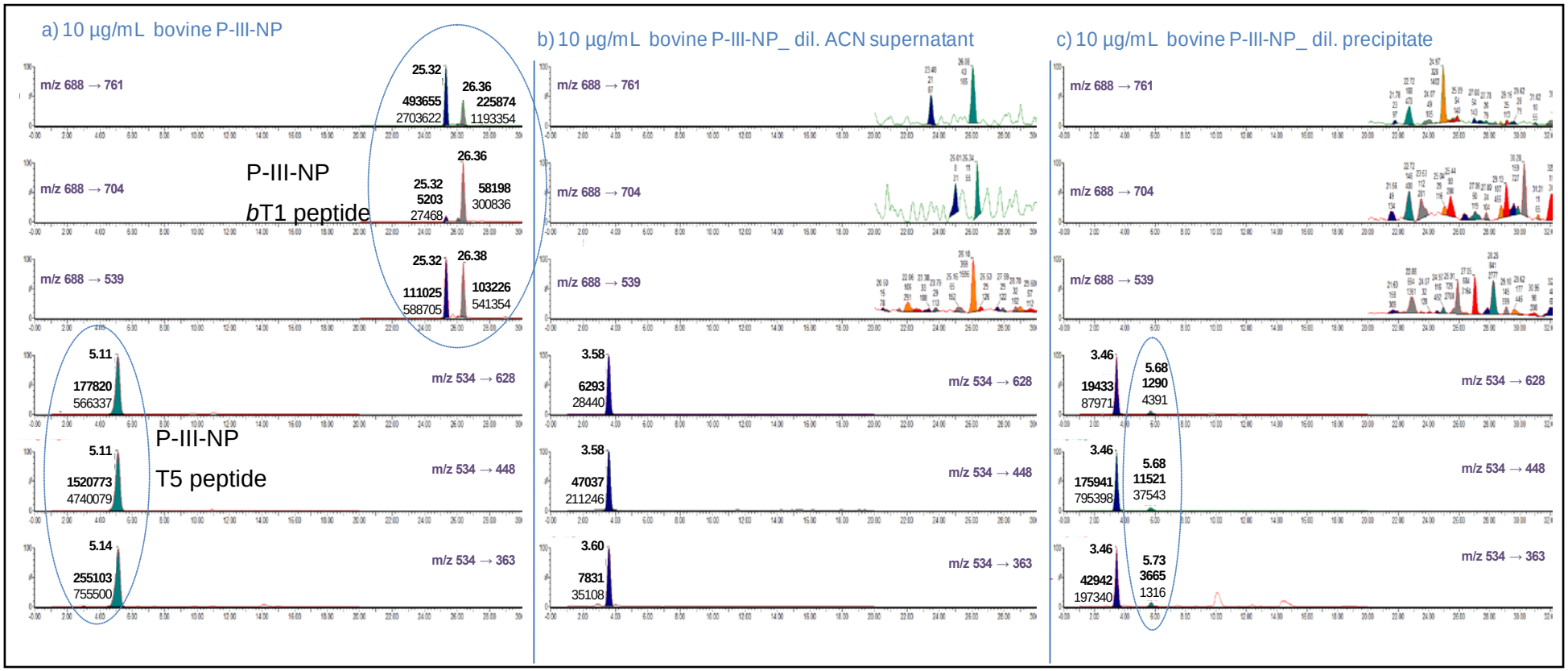


Figure 4.4-2: Chromatograms of a *b*P-III-NP sample (10 µg/mL) without HSA (a), and the acetonitrile and precipitate fractions collected from an acetonitrile protein precipitation (at pH 5.97) of diluted *b*P-III-NP (10 µg/mL) in 50 mg/mL human serum albumin (b and c, respectively), analysed by conventional-flow LC-MS after digesting overnight with trypsin. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *b*T1 and T5 peptides have been highlighted in a), and a small peak corresponding to T5 in c).

4.4.2 ***Separation of HSA and P-III-NP by MWCO filtration***

Separation of P-III-NP (42 kDa) and HSA (69 kDa) was achieved using MWCO filtration despite their similarity in size. Removal of the tertiary structure of P-III-NP by reduction and alkylation of the cysteine disulphide bond generates a monomeric protein (~14 kDa), that is one-third the size of its trimeric precursor. Similar treatment of HSA however, only results in the elongation of the protein which retains its MW of ~66 kDa. Thus, through the cleavage of the disuphide bonds in these proteins, a sufficient difference in MW is created that allows for their separation with a 30 kDA MWCO filter as is shown in Figure 4-9 (c) below, where the *b*T1 and T5 P-III-NP peptides were observed after a trypsin digestion of the filtrate.

Approximately 50 % of the *b*P-III-NP was recovered in the filtrate, compared to the control sample (a). Like ppt experiments, the recovery of proteins from MWCO filtration is not expected to be 100 %. Non-specific binding of proteins to the filter can occur. To reduce this the filter columns were passivated with a surfactant, Tween-20 (5 %), prior to use so as to block the non-specific binding sites. Loss of P-III-NP could also be due to protein being trapped within the HSA in the filter; to reduce this samples are diluted with solvent (ammonium bicarbonate solution) prior to loading on to the column.

With a reduction of *b*P-III-NP (100 ng/mL) in a similar matrix (50 mg/mL HSA), no *bT1* and T5 peptides were observed in the filtrate. This lack in observation of the peptides could be due to ion suppression caused by HSA peptides, as demonstrated in section 3.4.3, where the presence of HSA > 20 µg/mL suppresses the MS signal of the P-III-NP peptides. To enable analysis of serum P-III-NP in matrix (50 mg/mL HSA, equivalent to the highest expected serum basal concentration) it is essential that a HSA depletion efficiency > 99.6 % is achieved.

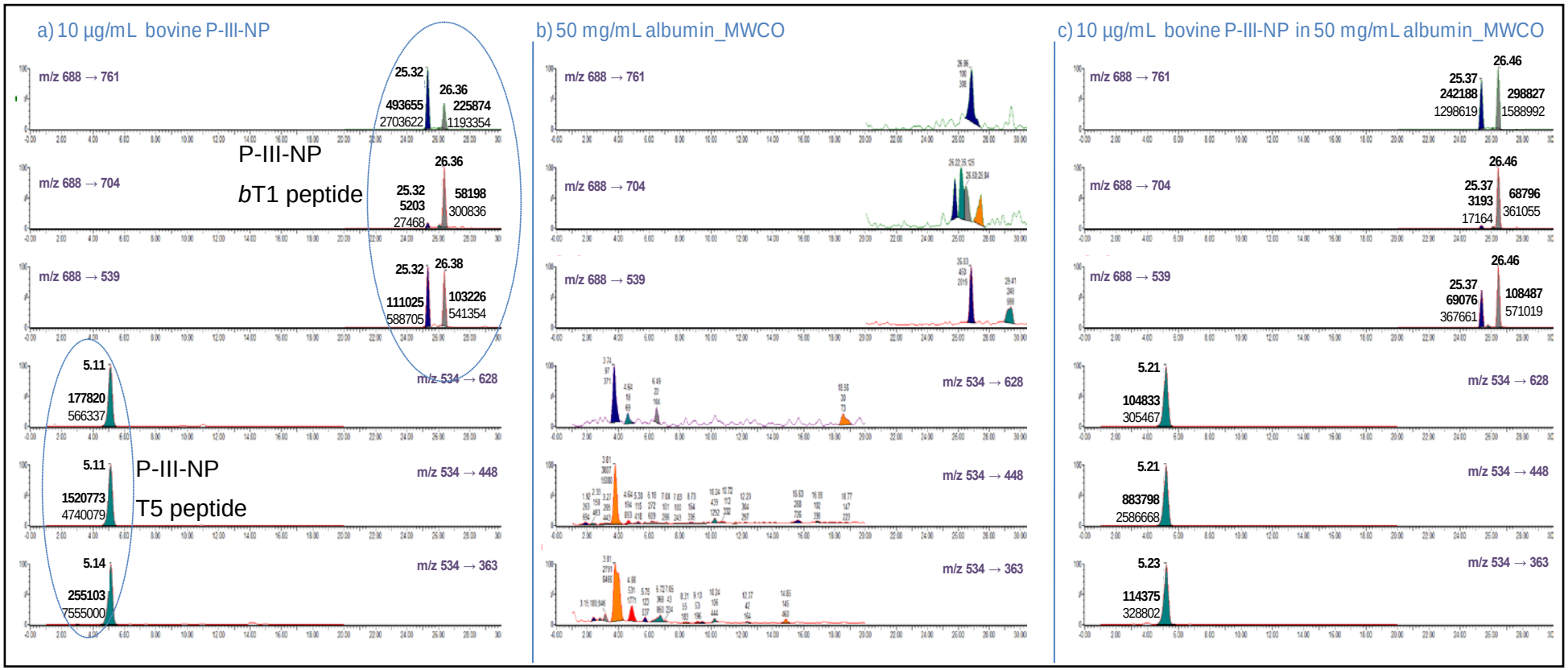


Figure 4.4-3: Chromatograms of a) *b*P-III-NP sample (10 µg/mL), b) MWCO filtered human serum albumin (50 mg/mL) and c) *b*P-III-NP (10 µg/mL) in 50 mg/mL human serum albumin, analysed by conventional-flow LC-MS after digesting overnight with trypsin. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *b*T1 and T5 peptides have been highlighted in a).

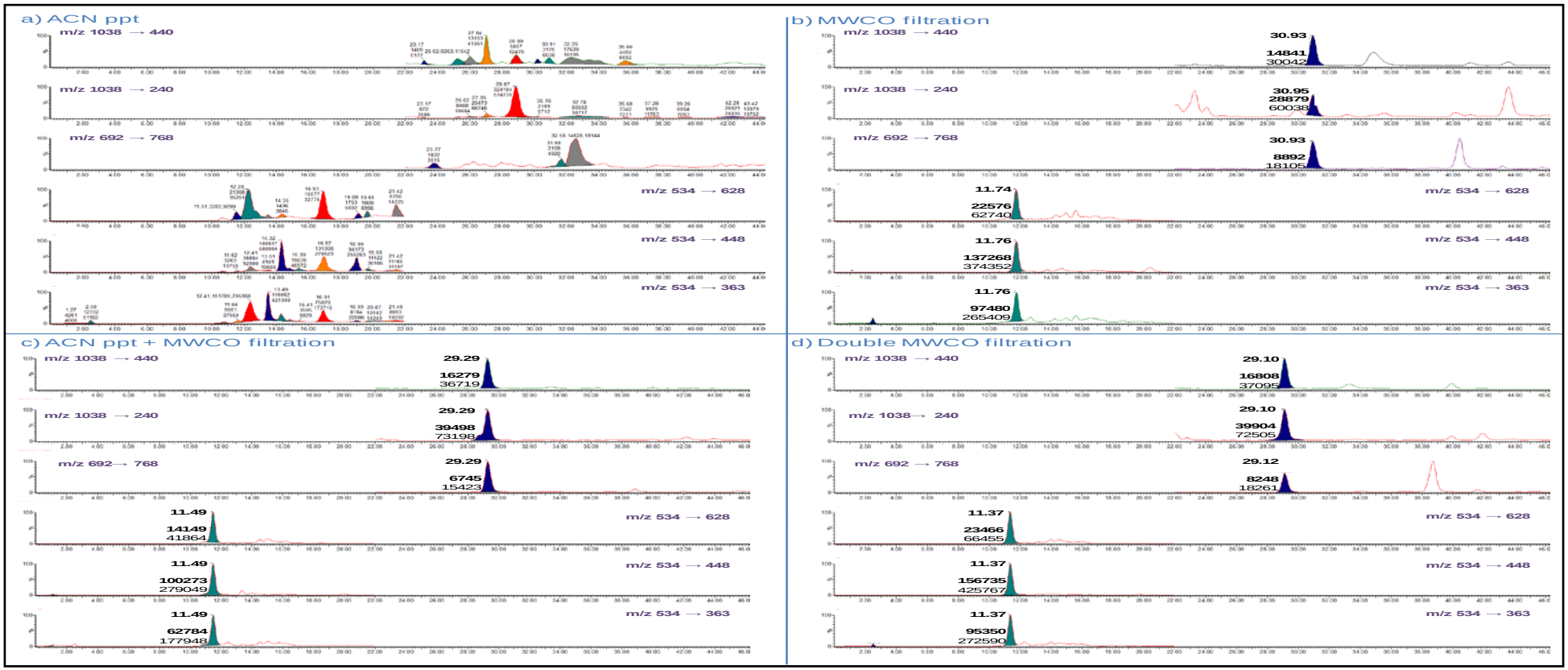


Figure 4.4-4: Chromatograms of P-III-NP *h*T1 and T5 peptides (100 pM, 10 µL) spiked in HSA depleted serum sample using acetonitrile protein precipitation (ACN ppt) and molecular weight cut-off (MWCO) filtration (30 kDa sieve) analysed by nano-flow LC-MS. Where the t_R of T5 and *h*T1 are 11.5±0.2 min and 30.1±1 min, respectively.

4.4.3 ***Efficiency of albumin depletion by ACN ppt and MWCO filtration***

So far we have established that separation of P-III-NP and HSA is necessary to analyse serum P-III-NP by LC-MS. Protein precipitation is unable to carry out this function, however, at high P-III-NP concentration (10 µg/mL) separation using MWCO filtration is possible. At lower concentrations (e.g. 100 ng/mL) P-III-NP peptides were not detectable in the MWCO filtrate of a HSA (50 mg/mL, 100 µL) sample. This is thought to be due to insufficient HSA removal that would cause ion suppression of the T1 and T5 peptides in the MS.

Table 4-5 shows the concentration of HSA remaining after MWCO and ppt of HSA (50 mg/mL) and serum (pooled) samples. MWCO and ppt on their own are insufficient in its removal of HSA. However, when combined (as either ppt then MWCO or double MWCO) the depletion efficiency of these techniques for HSA is drastically improved. Some of the proteins found in the depleted matrix are listed in Table 4-5 with a complete list shown in *Appendix I*. Data dependent analysis of the depleted serum samples show ALBU_H (HSA) present in all matrices, except the combined ppt and MWCO filtered sample. This observation does not correlate with that of the sample used to determine HSA concentrations by our targeted LC-MS methods. This, highlights the lack of reproducibility of the ppt and MWCO combination. Although sufficient depletion is achieved by this approach, the co-precipitation of P-III-NP and HSA limits its use for our analysis. This sample however is useful for determining the effects of other serum proteins on P-III-NP LC-MS analysis.

Spiked into depleted serum samples, *h*T1 and T5 (10 pM) peptides were observed in all matrices except that obtained by ppt (see Figure 4-10 below). Table 4-5 shows a large quantity of HSA remaining in the ppt depleted samples, which would result in the complete suppression of the P-III-NP peptides. Absence of other serum proteins in the HSA sample drastically improves the efficiency of depletion by this approach. This highlights the lack of specificity and completeness of acetonitrile precipitation. Post ppt of serum the residual proteins are sufficient to completely suppress *h*T1 and T5 in the

MS. Unfortunately the insufficient HSA depletion achieved by this method does not allow us to assess the effect of the other serum proteins. When combined with MWCO, the serum proteins remaining in the matrix do not affect the analysis of P-III-NP by LC-MS.

Table 4-5: Efficiency of HSA removal achieved by acetonitrile protein precipitation (ACN ppt) and molecular weight cut off (MWCO) filtration using a 30 kDa sieve. Some of the more abundant proteins remaining in depleted serum matrix (with a threshold > 80 %), identified using Mascot algorithm version 2.2.06 searching against the UniProt Swiss-Prot protein database are listed.

	Depleted serum albumin (µg/mL)		**Proteins identified in depleted serum matrix (peptide threshold > 80 %)**
	50 mg/mL HSA	Pooled human serum	
ACN ppt	40	1149	ALBU_H, APOA1, APOA2, HBA1, HBB, TRFE, CXCL7, APOA4
MWCO filter	30	80	ALBU_H, APOA1, APOA4
ACN ppt + MWCO filter	-	22	APOA1, APOA2, HBB, TRFE, CXCL7
Double MWCO filter	-	4	APO2, ALBU_H, APOA4

It is surprising to observe the *h*T1 and T5 peptides (10 pM) in the single MWCO depleted serum sample with signal intensities similar to the other matrices. Considering the early failure to separate *b*P-III-NP (100 ng/mL, 2.5 nM) from HSA (50 mg/mL) lack of complementarity of these observations could be due to insufficient reproducibility of MWCO filters, trapping of *b*P-III-NP in the HSA that is cut-off by the filter, or losses caused by sample handling and non-specific binding. Like ppt, MWCO separation appears to be more efficient in the HSA sample than serum, however for this approach the levels of depletion of the two matrices are more comparable. The MWCO filter should trap all proteins > 30 kDa; nevertheless ALUB_H (66 kDa) and ALPOA1 (45 kDa) are among the proteins observed in the filtrate. The observance of these proteins could be due to the capacity of the filter being exceeded. However, this is unlikely as with double MWCO filtration HSA remains in the filtrate. Thus, irrespective of size, some proteins (> 30 kDa) will escape the filter and the efficiency of MWCO will always be < 100 %. Sufficient depletion is achieved using double MWCO filtration,

which is capable of separating P-III-NP and HSA, thus this approach is further investigated for use in the preparation of serum for P-III-NP LC-MS analysis.

4.4.4 *Recovery of P-III-NP from double MWCO filtration*

With double MWCO separation of *b*P-III-NP in 50 mg/mL HSA (100 ng/mL, 100 µL) a 12 % recovery of P-III-NP was observed (n > 10). Sporadically recoveries > 50 % have been observed; this could be due to a lack in reproducibility of albumin depletion achieved by the MWCO filter column. MWCO filters are designed to capture higher MW proteins than the filter cut-off; hence, in its current use we are inverting its primary function for which the efficiency of this approach is not defined by the manufacturer. To improve recovery from the filter, potential MWCO carrier proteins (tetracosactide, substance P and recombinant human insulin) which should limit non-specific binding of P-III-NP to the filter to allow for passage into the filtrate were investigated.

4.4.4.1 Carrier proteins

Based on the size of P-III-NP (42 kDa) it is not expected to bind to HSA (66 kDa) in serum; hence, HSA is not a carrier for P-III-NP. At low concentrations, similar to that of basal serum P-III-NP (5 ng/mL, 125 pM), proteins are prone to sample handling losses especially by non-specific binding. Thus, the removal of albumin during MWCO filtration can increase the loss of P-III-NP during sample handling. To avoid this, the presence of small but highly concentrated proteins (peptides) able to pass through the filter and bind to non-specific sites within the filter and also bind the sample capture vessels were investigated. This was determined in terms of their effect on minimising P-III-NP loss.

Substance P, a 1346 Da neurotransmitter protein, has the aa sequence **R**P**K**PQQFFGLM. This peptide contains two potential cleavage sites for trypsin (highlighted in bold), however due to the close proximity of Pro to these sites there is likely to be very little or no cleavage from the endoprotease. Interference from a trypsin digest of the peptide was investigated with *h*T1 and T5 peptides (10 pM); chromatograms are shown below in Figure 4-11. For T5 some suppression is observed

in the *m/z* 534 to 448 and 363 transitions, however no interference was observed for the *h*T1 peaks. In full scan mode, Substance P was seen to elute at the end of the run, hence, it should not be the cause of the suppression observed.

Tetracosactide (adrenocorticotropic hormone (ACTH) 1-24) is a 2933 Da protein with the aa sequence SYSMEHF**R**WG**K**PVG**KKRR**PV**K**VYP. Several potential tryptic cleavage points exists in this peptide, however, due to the close proximity of Pro or multiple neighbouring cleavage sites the probability of cleavage is affected. Analysis of P-III-NP peptides with carrier show some interference for T5 and none for *h*T1. Like substance P, the intact protein elutes at the end of the gradient in full scan; however, no investigation into the t_R for potential tryptic peptides was done.

Human insulin, a 6 kDa protein, has two potential tryptic cleavage sites within its b-chain which has the aa sequence FVNQHLCGDHLVEALYLVCGE**R**GFFYTP**K**T. The cleavage closer to the c-terminus of the protein is unlikely, due to the neighbouring Pro in the P2 secondary trypsin binding site (relative to K), hence cleavage should yield peptides FVNQHLCGDHLVEALYLVCGER and GFFYTPKT. Like the other proteins, some suppression of T5 is observed, despite the intact insulin protein and its tryptic peptides eluting at the end of the LC gradient. Insulin shows the least (although insignificant) interference for the *h*T1 and T5 peptides, hence, it is selected as our model protein to investigate its effect as carrier for P-III-NP during MWCO filtration.

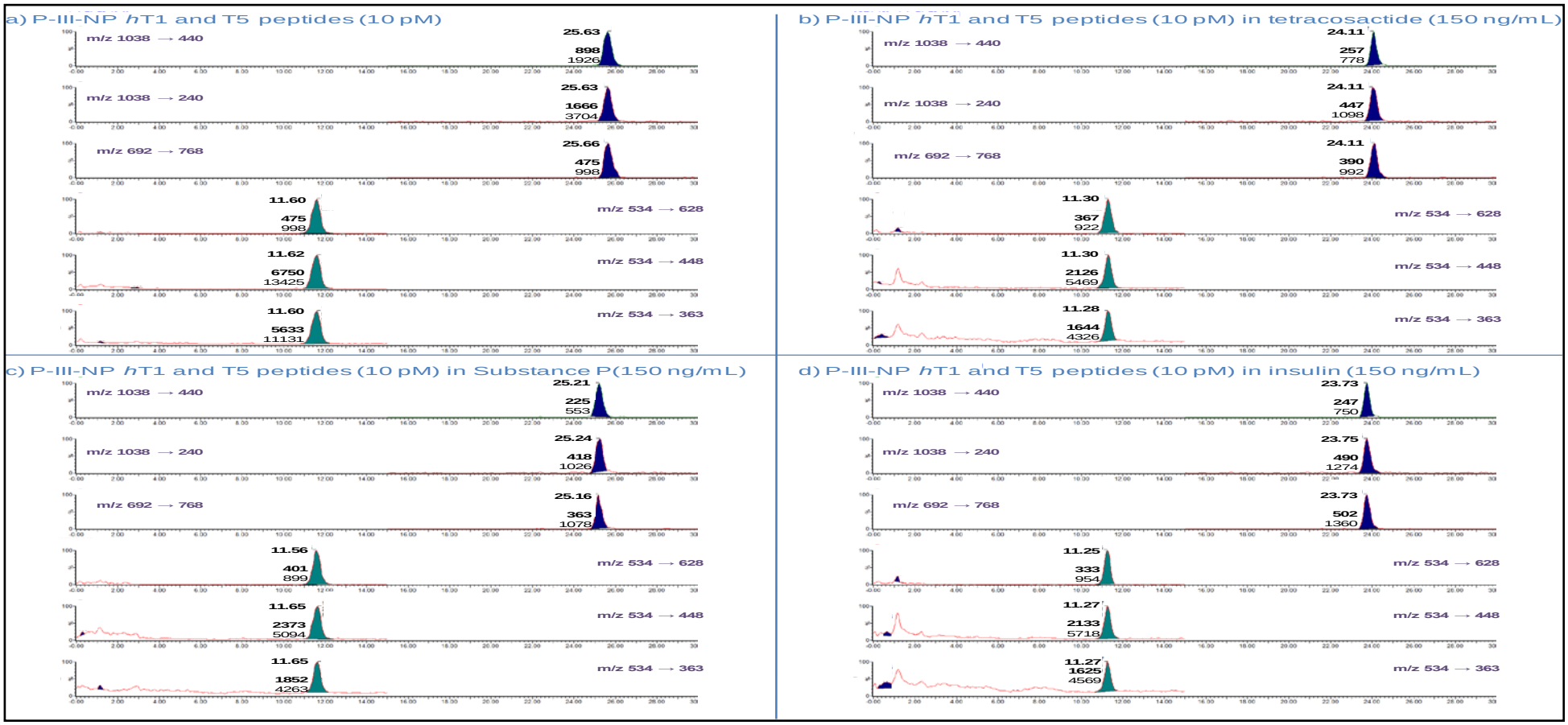


Figure 4.4-5: Chromatograms of P-III-NP hT1 and T5 peptides (10 pM, 5 µL) analysed by nano-flow LC-MS with and without potential carrier proteins tetracosactide (a), Substance P (c), and insulin (d). Where the t_R of T5 and *h*T1 are 11.5±0.2 min and 24.1±1 min, respectively.

4.4.4.2 Assessment of recovery using a carrier protein

Overall, no improvement in the recovery of P-III-NP separated from HSA by double MWCO filtration was achieved with insulin acting as a carrier (see Table 4-6 below). Compared to the control sample (Tween-20 passivated), no improvement in the MS signal for the T5 peptide was observed for the Tween-20 passivated and insulin (in-solution) sample. For *b*T1, the observed recovery is misleading, as observed signals are likely from interferences. Extensive suppression in one of the replicates was observed for *b*T1 and the %RSD for T5 peak was significantly high (> 51 %).

For the insulin passivated MWCO filter sample, complete suppression of *b*T1 was observed in both samples. A small decrease in T5 recovery was observed with similar %RSD, compared to the control sample. For the insulin passivated and insulin (in-solution) sample, a decrease in recovery was observed for the T5 and *b*T1 peptides. Of the systems investigated, this combination gave the lowest recovery for P-III-NP, especially for the T5 peptide.

The %RSD for the duplicate samples are sporadic in all MWCO filtered combinations, which is not surprising as the reproducibility of ultrafiltration has been reported as being poor [179, 180]. This is likely associated with interferences in the matrix. Despite double MWCO filtration being capable of separating HSA from P-III-NP, interferences from the matrix and low recovery makes this system difficult to work with, as such this approach is not further investigated.

Table 4-6: Recovery (peak area) of trypsin digested *b*P-III-NP (100 ng/mL, 100 µL) separated from HSA (50 mg/mL) using double MWCO filtration. The effect of insulin as potential carrier protein is included. Values are displayed as an average (n=2) except for those in *blue*, where n=1.

		100 ng/mL Control	Tween-20 passivated (avg)	% Recovery	% RSD	Tween-20 passivated + Insulin (avg)	% Recovery	% RSD	Insulin passivated (avg)	% Recovery	% RSD	Insulin passivated + Insulin (avg)	% Recovery	% RSD
Area	*m/z*													
T5	534 -> 628	103 314	15 254	15	0.9	5 490	5	76.0	11 746	11	13.1	3642	4	9.9
	534 -> 448	639 927	99 880	16	5.4	30 479	5	70.2	72 858	11	6.4	24 144	4	18.4
	534 -> 363	470 229	72 211	15	10.3	28 066	6	51.6	52 297	11	12.1	15 177	3	5.5
*b*T1	688 -> 761	256 979	7 202	3	2.9	*4 781*	2	-	-	-	-	2 085	1	33.3
	688 -> 440	93 832	5 179	6	33.0	*-*	-	-	-	-	-	6 363	7	9.1
	688 -> 369	72 594	4 698	6	0.5	*6 004*	8	-	-	-	-	3 339	5	16.7
	688 -> 240	78 184	3 640	5	35.8	*14 748*	19	-	-	-	-	3 396	4	9.6
Height														
T5	534 -> 628	404 054	51 029	13	5.1	17 585	4	71.9	33 973	8	10.4	12 048	3	4.7
	534 -> 448	2 514 786	333 595	13	5.9	99 235	4	94.6	217 821	9	9.1	74 027	3	4.0
	534 -> 363	1 834 109	244 522	13	6.0	85 146	5	69.7	158 873	9	9.9	51 796	3	3.6
*b*T1	688 -> 761	621 173	15 770	3	6.9	*8 920*	1	-	-	-	-	5 423	1	24.6
	688 -> 440	93 832	7 689	8	26.4	*4 207*	4	-	-	-	-	8 968	10	34.6
	688 -> 369	72 594	5 559	8	22.3	*12 098*	17	-	-	-	-	5 741	8	9.2
	688 -> 240	18 8134	6 349	3	14.6	*26 820*	14	-	-	-	-	7 638	4	7.5

4.4.5 ***Immunocapture of P-III-NP***

Separation of P-III-NP and HSA was achieved using immunocapture with mouse monoclonal P-III-NP antibody (CisBio Bioassays). Table 4-7 shows the recovery of *b*P-III-NP for 5 ng/mL and 500 ng/mL samples, which was > 70 % for all peptide MS transitions. The %RSD for all sample pairs was ≤ 25; a drastic improvement to that observed using MWCO filters. The calculated albumin depletion of this method using HSA (50 mg/mL, 100 µL) was 99.99 %; with ~60 ng/mL HSA remaining in the sample after depletion.

Table 4-7: Recovery (peak area) of *b*P-III-NP separated from HSA (50 mg/mL) by immunocapture. Values are displayed as an average of n=2.

	m/z	*b*P-III-NP control (5 ng/mL)	Immuno-capture *b*P-IIII-NP (5 ng/mL) sample	% Recovery	%RSD	*b*P-III-NP control (50 ng/mL)	Immuno-capture *b*P-IIII-NP (50 ng/mL) sample	% Recovery	%RSD
T5	534 -> 628	8 585	20 261	236	18	126 529	139 329	110	7
	534 -> 448	45 120	133 603	296	10	716 010	933 004	130	8
	534 -> 363	35 456	91 133	257	13	521 067	640 381	123	8
*b*T1	687 -> 761	24 778	22 670	91	15	294 541	217 183	74	22
	687 -> 440	8 895	8 349	94	10	97 713	72 513	74	23
	687 -> 369	6 014	7 599	126	1	74 059	60 332	81	23
	687 -> 240	6 631	7 803	118	11	84 634	73 096	86	25

For the pseudo-serum basal *b*P-III-NP sample (5 ng/mL) recoveries much greater than 100 % for the T5 peptide was observed despite the identical volumes (100 µL) being analysed. This disparity in the results could be due to non-specific binding loss occurring in the control sample which was analysed in 0.01 % HSA to circumvent matrix effects. In the sample separated by immunocapture, a significantly higher concentration of HSA is present that would reduce this loss. For the *b*T1 peptide ~100 % recovery is observed; in reality however, this will be much decreased following the pattern of T5 (the more MS sensitive of the two peptides). Like the 5 ng/mL sample, the recovery for *b*T1 is < T5, as expected. The degree of loss is expected to be less between standard and sample, thus should be more realistic in its estimation of loss.

For the 50 ng/mL *b*P-III-NP sample the recoveries are slightly greater than 100 % for the T5 peptide. At higher protein concentration the loss is expected to be reduced, as has been observed. However in the absence of high HSA concentrations, loss is still expected. The better correlation of data at this concentration is more likely a reflection of similarity in the analyte loss to non-specific binding in the control to the losses in the capture of P-III-NP by the antibody for the pseudo-serum sample.

The observed recovery and albumin depletion of this approach sufficiently satisfies the criteria required for the sample preparation of serum containing *h*P-III-NP (at basal concentrations) for LC-MS analysis using the developed methods. Application to a pooled serum sample, analysed by micro-flow and nano-flow LC-MS after trypsin digestion are shown in Figure 4-12, Figure 4-13, respectively. The *h*T1 and T5 peptides are qualitatively identified in the sample by both analytical methods. Interference (from HSA) for the *h*T1 peptide is observed (see Figure 4-14 below) as the ion ratio and peak intensity of this peptide is anomalous compared to the standard. A peak corresponding to *h*T1 is seen in the immunocapture HSA sample; samples fortified with pure *h*T1 peptide show an increase in the peak area. Hence, further development of the LC-MS for *h*T1 is required, especially if it is to be used quantitatively to determine P-III-NP. For the T5 peptide, however, the ion ratios (tolerance window of relative abundance ± 20 %) and peak intensity (signal to noise > 3:1) are within the expected range for a basal serum sample [181]. Also, no interference is observed in the immunocaptured HSA sample (Figure 4-14). Targeting the T5 peptide a semi-quantitative approach was used to estimate a P-III-NP concentration of 2 ng/mL (50 pM) using a 200 µL volume for this pooled serum sample. By immunoassay (Orion Diagnostica) this sample was determined to have an average P-III-NP concentration of 3.2 ng/mL.

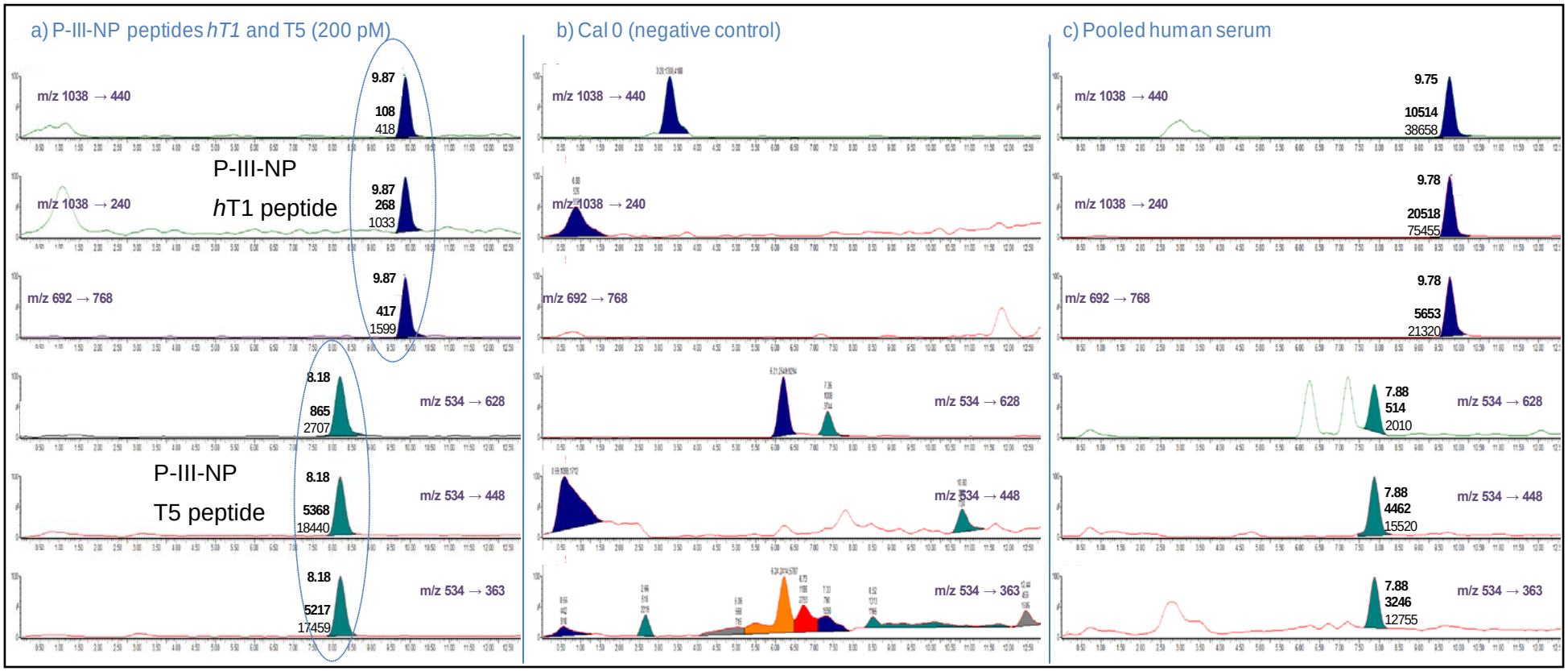


Figure 4.4-6: Chromatograms of P-III-NP *h*T1 and T5 peptides (100 pM) (a) and immunocaptured P-III-NP from Cal 0 (containing bovine proteins, preservatives and a yellow orange dye, CisBio Bioassays) a pooled human serum samples (200 µL) (b and c, respectively), analysed by micro-flow LC-MS after digesting overnight with trypsin. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *h*T1 and T5 peptides have been highlighted in a).

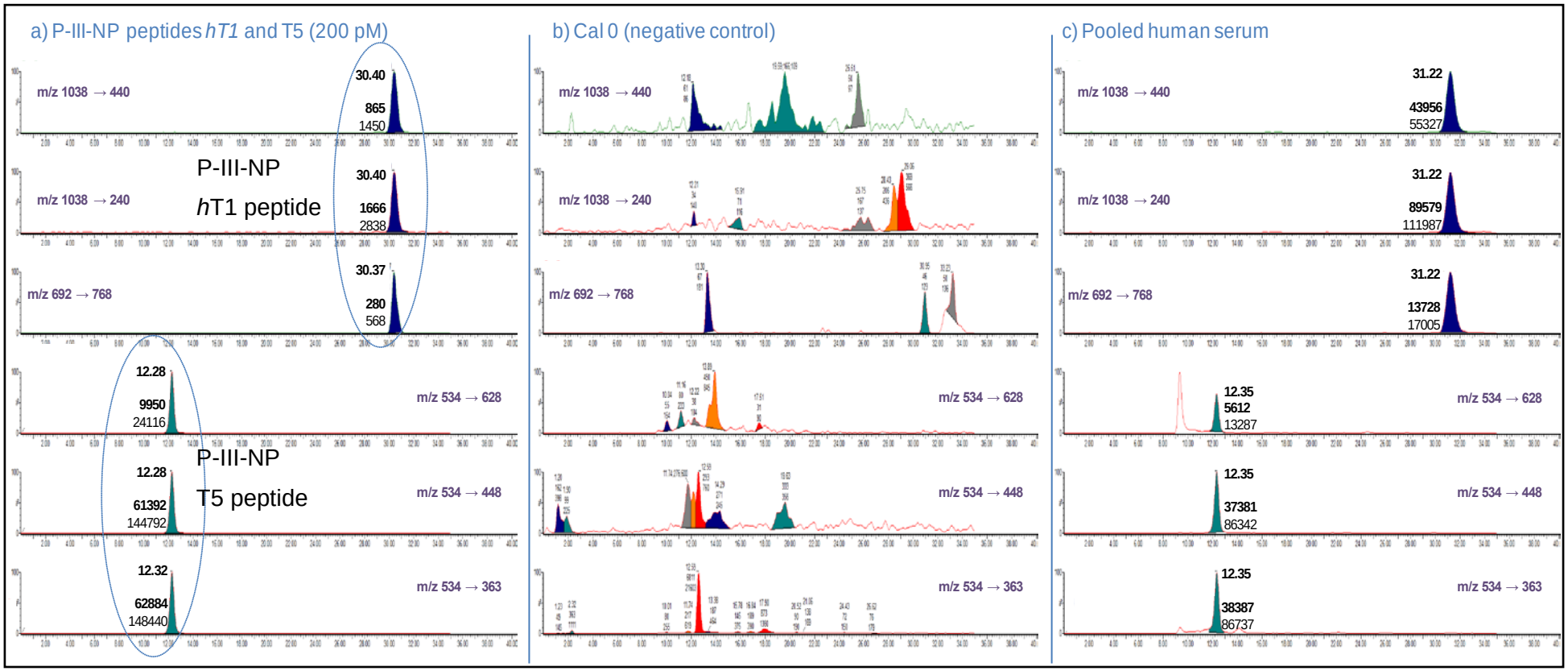


Figure 4.4-7: Chromatograms of P-III-NP *h*T1 and T5 peptides (100 pM) (a) and trypsin digested immunocaptured P-III-NP from Cal 0 (containing bovine proteins, preservatives and a yellow orange dye, 200 µL, CisBio Bioassays) and a pooled human serum sample (200 µL) (b and c, respectively), analysed by nano-flow LC-MS after digesting overnight with trypsin. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *h*T1 and T5 peptides have been highlighted in a).

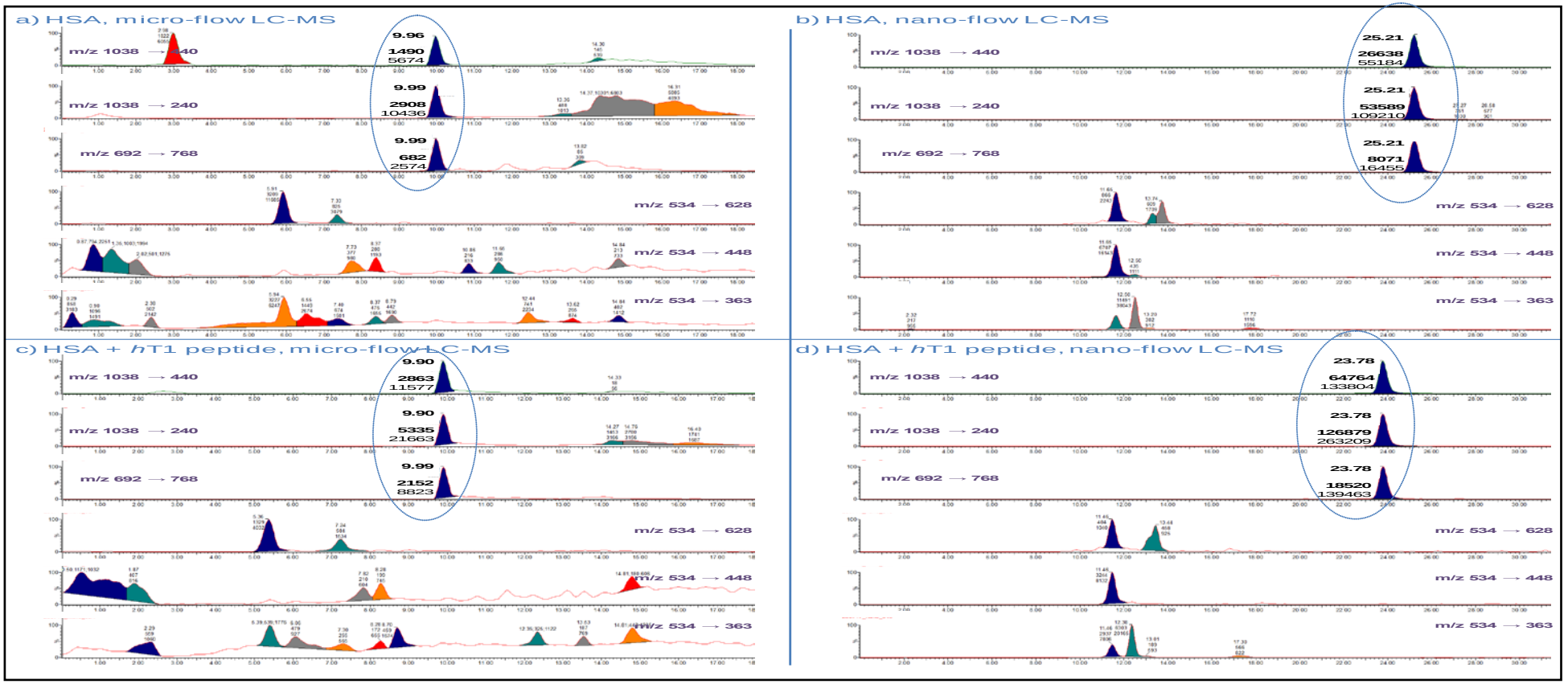


Figure 4.4-8: Chromatograms of trypsin digested immunocaptured HSA (50 mg/mL, 200 µL) and fortified *h*T1 samples analysed by micro-flow (a and c, respectively) and nano-flow (b and d, respectively) LC-MS. Retention time, peak height and peak area (in bold) are shown for each peak. The eluting region for *h*T1is highlighted.

4.5 **Summary**

P-III-NP can be separated from HSA in serum using double MWCO filtration or targeted P-III-NP immunocapture methods with albumin depletion efficiencies > 99.6 %. Resulting from co-precipitation of the two proteins, separation cannot be achieved using protein precipitation. Double MWCO filtration gives a low recovery of P-III-NP (even in the presence of a carrier protein) and the reproducibility of the filter is questionable, hence this approach was not further investigated. Immunocapture gives a reproducible recovery of P-III-NP > 70 %. Application of this method to a pooled human serum sample containing P-III-NP at basal concentrations gave an estimated concentration of 2 ng/mL (50 pM) when analysed by the developed micro- and nano-flow LC-MS methods. So far suitable sample preparation and LC-MS methods have been developed which enable the detection of serum *h*P-III-NP. However to associate traceability to these methods and eventually enable their use for absolute quantification, suitable reference material for *h*T1 and T5 or intact *h*P-III-NP is needed. Thus, in chapter 5 the initial steps for synthesis of appropriate peptide standards are discussed.

Chapter 5 *: An approach to obtaining peptide standards for the quantification of P-III-NP*

5.1 Overview of the chapter

In the previous chapter, *h*P-III-NP was extracted and analysed by LC-MS from human serum to give an approximate concentration of 2 ng/mL (50 pM) against the T5 peptide standard. For absolute quantification of P-III-NP, suitable reference peptides or intact protein are needed. In the absence of reference material for *h*P-III-NP, synthesis of ^{13}C and ^{15}N isotopically labelled and unlabelled peptides for *h*T1 and T5 was initiated. Once synthesised, absolute quantification of the acid hydrolysed aa subunits of each peptide is to be quantified by targeted LC-MS approaches. With sufficient synthesis and quantification, these peptides are to be used as internal reference standards for the absolute quantification of serum *h*P-III-NP, using the developed digest P-III-NP LC-MS approach described in Chapter 3.

5.2 Introduction

In MS-based quantitative analysis, the unknown concentration of a target sample analyte is determined by comparison with the absolute MS signals of a set of known standards. To achieve accurate concentrations, reproducible analyte measurements are necessary; this is achieved with minimum variance in sample extraction, preparation, and analytical methods. Stable-isotopic standards (with similar chemical properties to the target analyte) can be incorporated in the sample workflow to normalise these variances, however the earlier they are included the more effective they are at improving the reproducibility of methods [182, 183].

The absence of reference material for *h*P-III-NP makes application of protein labelling impossible. However, if the proposed digest approach to measurement is adopted, peptide labelling becomes appropriate and heavy isotope labelled (typically ^{13}C and/or ^{15}N) variants of target surrogate peptides, *h*T1 and T5, can be incorporated into the sample workflow. Ideally both the labelled and unlabelled peptide variants should co-

elute and display similar MS/MS fragmentation patterns, despite the presence of the isotopic labelling on the heavy-variant; which will associate a difference in mass to the precursor and product ions. With the incorporation of these heavy labelled standards in the sample workflow directly after protein digestion, the co-analysis of both analytes should correct any analytical variances associated with LC-MS methods as well as protect against false bias originating from matrix interference.

5.2.1 ***Peptide synthesis***

Peptides fragments of P-III-NP are not commercially available hence they require synthesis. Mimicking *in vivo* protein production, peptides (and proteins) can be chemically synthesised by the sequential coupling of the respective amino acids in its sequence. Solution and solid phase approaches have been used for peptide and protein synthesis, though the latter approach is preferred based on its simplicity, speed, efficiency and amenability to automation [184, 185].

During synthesis, peptides are elongated from their carboxyl- to amino-terminus to facilitate the control of sequence variation with the sequential addition of reversibly-protected α-amino group residues. To control the location of peptide bond formation during synthesis, aa residues containing additional reactive side groups are also reversibly blocked. In solid-phase peptide synthesis (SPPS), insoluble porous resins (normally containing polystyrene) are used to anchor the growing peptide. The associated insolubility of the peptide, as a result of its attachment to the resin, facilitates the continuous removal of the impurities and excess aa in the sample mixture throughout peptide synthesis. For the coupling of amino- and carboxyl-ends of two aa residues to occur, the carbonyl- group needs to be activated, this is achieved with the inclusion of coupling agents such as DIC, HATU and DIPEA (refer to list of abbreviations or materials and methods below for full compound names) in the reaction mixture.

For the α-amino group, the most commonly used protecting groups are tert-butoxy-carbonyl (Boc) and fluoren-9-ylmethyloxycarbonyl (Fmoc). Boc protecting

groups are sensitive to acids such as TFA, whilst Fmoc groups are sensitive to bases like piperidine. Thus, based on their reactivity and effective removal from aa residues, Fmoc strategies are preferred in routine peptide synthesis. The milder conditions required for Fmoc removal allows orthogonal deprotection of aa residues during synthesis, without disruption of neither the peptide-solid support link nor any side chain protecting groups (e.g. Trt, Pbf, O^tBu and Boc), which are normally acid sensitive. Fmoc decoupling can be monitored by UV spectroscopy, where the absorbance of dibenzoflulvene-piperidine complexes (formed between piperidine and the cleaved Fmoc group) at 301 nm are used to indicate successful cleavage of the Fmoc protecting group during synthesis [185]. Alternatively decoupling of α-amino protecting groups can be monitored using several qualitative tests (e.g. Kaiser and picrylsulphonic acid tests) that are sensitive to primary amino groups [186, 187].

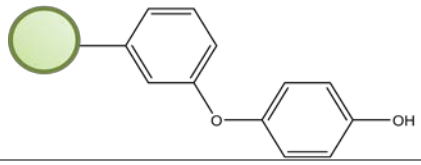
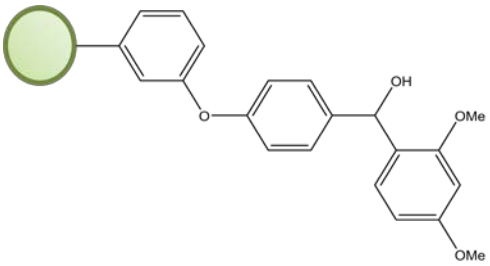
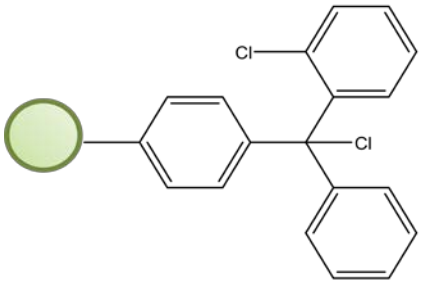
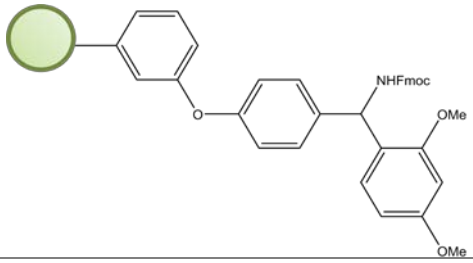
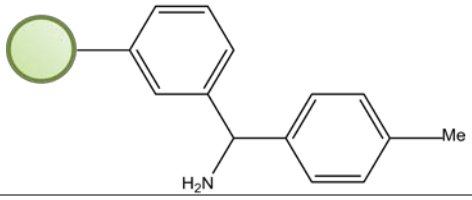
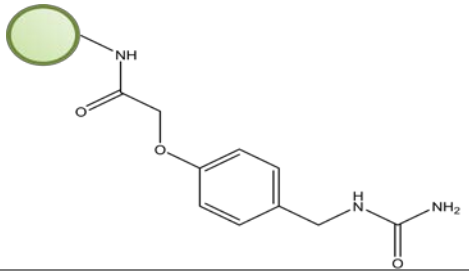
For SPPS, the core of the resin used is normally made of polystyrene (with 1-2 % divinylbenzene), and are usually small and spherical in shape. The swelling of the polystyrene resins is important for synthesis, in that expansion of the polymer affects the diffusion and accessibility of reagents to the core. These resins are ideally swelled to their maximum in organic solvents, such as dimethylformamide (DMF) ethanol or acetonitrile, before commencing synthesis. Attachment of peptides to the polystyrene core of resins is achieved through linkers on their surface. The density of linkers attached dictates the loading capacity of the resin, which defines the quantity of peptide that can be synthesised. Manufacturers normally express loading capacities in mmol/g, where values represent the concentration of linker per gram of resin. A variety of linkers have been developed for polystyrene based resins; some are listed in Table 5-1 with their final cleavage conditions and synthesised peptide carboxyl-terminal functional groups [185]. Most linkers are cleaved in acidic conditions and produce peptides with either an acid (COOH) or amide ($CONH_2$) functional group at the peptide carboxyl-terminus. Resins that are preloaded with the first aa are also available. Use of

these resins for SPPS will increase the efficiency of yield, where the first aa coupling is a limiting factor.

5.2.2 ***Validation of peptide standards***

For use as standards in analytical methods, peptides need to possess chemical purity of > 95 % and be accurately quantified, to ensure the quality of their measurements meets the specification of the assay (i.e. specificity, precision, bias, lower limits of quantification) [182, 188]. Purification is achieved by the fractionation of crude peptide samples, and can be monitored by elemental analysis, Karl Fisher analysis (for water content), LC-UV or NMR spectroscopy. Without purification, quantification of peptides is impossible, as measurements will most certainly include impurities. Whilst gravimetric methods are applicable, aa analysis of acid hydrolysed peptides by LC-MS is a better approach to peptide quantification. Here single aa concentrations of a known amount (accurately measured weight) of hydrolysed peptide is used to determine the net peptide content. By the multiplexed analysis of different aa residues within the peptide sequence, cross-comparison of determined concentrations can also be used to support the purity of the peptide sequences, which in addition to peptide sequence, MS analysis (preferably MS/MS) confirms the success of synthesis.

Table 5-1: List of some of the resins available for SPPS, for which the cleavage conditions and the carboxyl-terminus functional group of the synthesised peptide are given.

Resin name	Resin structure	Cleavage conditions	Peptide produced
Wang		90-95 % TFA in dichlormethane (1-2 h)	Acid
Rink		1-5 % TFA in dichloromethane (5-15 min)	Acid
2-chlorotriyl chloride		1-5 % TFA in dichloromethane (1 min)	Acid
Rink amide		150 % TFA in dichloromethane (1 hr)	Amide
MHBA		Hydrogen flouride 0 °C (1 h)	Amide
PEGA-BAL		TFA-TFMSA (19:1)	Acid

5.3 **Materials and Methods**

The source of materials where unmentioned can be found in section 2.3, section 3.3 and section 4.3. Fmoc-Hyp(O^{t}Bu)-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Asp(O^{t}Bu)-OH, Fmoc-Ser(^{t}Bu)-OH, Fmoc-Glu(O^{t}Bu)-OH, L-pyroglutamatic acid, amino acid standard mix (analytical reagent, cat. no. AAS18, containing all natural aa except Asn and Gln), ethyl (hydroxyimino)cyanoacetate (97 %. Oxyma), 2,2′-(ethylenedioxy) diethanethiol (95 %. EDT), *N,N′*-Diisopropylcarbodiimide (98 %. DIC), 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluoro-phosphate (97 %. HATU), dimethylformamide (99.8 %, anhydrous. DMF) and piperidine (99 %, reagent plus®) were purchased from Sigma Aldrich (St. Louis, MO, USA). Fmoc-L-ala-OH, Fmoc-Tyr(^{t}Bu)-OH, Fmoc-Gln(trt)-OH, Fmoc-Leu-OH, Fmoc-His(trt)-OH, Fmoc-Cys(trt)-OH, and Fmoc-Val-OH were purchased from Hong Kong PE Biosciences Ltd (Shanghai, China). L-Ile-N-Fmoc ($^{13}C_6$, 97-99 %; ^{15}N 97-99 %), L-Arg:HCL ($^{13}C_6$, 99%) and L-Leu ($^{13}C_6$, 99%) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA, USA). Picrylsulfonic acid was purchased from Tokyo Chemical Industry Company Ltd. (Tokyo, Japan). Methanol, dichloromethane (DCM) and diethyl ether were purchased from Fisher Scientific (Loughborough, UK). *N,N*-Diisopropylethylamine trihydrofluoride (95 %) was purchased from Acros Organics™ (DIPEA. Fisher Scientific, Loughborough, UK). Phenol (ACS, Reag. Ph Eur) was purchased from Merck Millipore (Darmstadt, Germany). Triisopropylsilane, (98 %. TIPs) and thioanisole (99 %) were purchased from Alfa Aesar™ (Heysham, England).

5.3.1 ***Heavy labelled peptide development***

Using the optimised *h*T1 and T5 SRM transitions detailed in section 3.3.1.1, heavy labelled peptide variants were designed for each peptide to ensure that in all the monitored transition-fragments an isotopic $^{13}C/^{15}N$ aa residue was incorporated. The designed isotopic-labelled heavy peptide sequences are shown below:

Heavy *h*T1: Q_{-NH3}***-[^{13}C/^{15}N-Q]-***EAVEGGCSHLG***-[^{13}C/^{15}N-Q]-***SYADR

Heavy T5: GD(P_{+OH})GP(P_{+OH})G-***[^{13}C/^{15}N-I]***-(P_{+OH})GR

5.3.1.1 F-moc peptide synthesis

Peptides for *h*T1, T5 and heavy T5 were manually synthesised by Fmoc-solid phase peptide synthesis (SPPS). For the T5 peptides a Fmoc-Arg(Boc)-Wang resin (100-200 mesh) with a load capacity of 0.6 mmol/g (Bachem, Bubendorf, Switzerland) was used. For *h*T1 a Fmoc-Arg(Pbf)-NovaSyn®TGA resin with a load capacity of 0.18 mmol/g (Merck Millipore) was used. Each peptide was prepared using the following protocol.

1. The resin (200 mg) was swollen in DMF in a fritted polypropylene tube for 30 min. The resin was then washed with DMF, methanol and DCM before removing a small amount (< 0.1 mg) to which picrylsulphonic acid (20 µL) was added to check for Fmoc protection of the linked Arg with the observance of a negative amine reaction (no colour change to resin).
2. The Fmoc protecting group was removed by treating the resin with piperidine in DMF (20 % v/v) for 30 min. Subsequently, the piperidine was removed by filtering under vacuum before washing the resin with DMF, methanol and DCM, to remove impurities and excess reagent. Once washed, a small amount of resin (< 0.1 mg) was removed and tested with picrylsulphonic acid for a positive amine reaction (orange colour change in resin) to ensure Fmoc removal.
3. For each aa coupling (except ^{13}C/^{15}N aa and pyroE residues), a four equivalent (excess) mixture of the Fmoc-aa, DIC and oxyma (1:1:1 molar ratio) in DMF was prepared using the weights and volumes shown in Table 5-1, Table 5-2 and Table 5-3 below. For ^{13}C/^{15}N-Ile a three equivalent mixture was prepared. For pyroE a ten equivalent mixture of Fmoc-aa, HATU and DIPEA (1:1:1 molar ratio) in DMF was prepared. (Coupling mixtures were made no more than 2 min before adding to the resin).
4. After Fmoc removal, the resin was washed with DMF before adding the coupling mixture of the first aa of the peptide being synthesised. The resin was left to react with the coupling mixture for 2 h, after which excess reagents and

impurities (including by-products) were removed by washing with DMF, methanol and DCM. A small amount of resin (< 0.1 mg) was removed and tested with picrylsulphonic acid to check for Fmoc protection (indicative of successful aa coupling).

5. The Fmoc protecting group was then removed by repeating step 2.
6. Subsequent aa were added to the sequence by repeating steps 3-5, until the sequence was complete. Samples were dried overnight under vacuum before decoupling.

Once the peptides were synthesised, they were removed from the resin by incubating (3 h) in a cleavage solution (4 mL) containing TFA, water, phenol, TIPs, thioanisole and EDT (90:5:5:2.5:5:2.5 molar ratio). The cleaved peptide was then filtered into a clean Falcon tube under vacuum (residual peptide was washed off the beads with TFA (~1 mL)). The filtrate was then concentrated to a volume < 500 µL by heating (~40 °C) under nitrogen before precipitating the peptide with ice-cold diethyl ether (~4 °C). Samples were then centrifuged at 4000 rpm for 10 min and the supernatant discarded. The precipitate was washed 2-3 times with diethyl ether to remove impurities, before the remaining pellet was freeze-dried for 24 h. The lyophilised product was stored at 4 °C.

Table 5-2: Components of the coupling mixture for each aa used in the Fmoc-SPPS synthesis of the T5 peptide.

F-moc AA residue	1 Gly	2 Hyp	3 Ile	4 Gly	5 Hyp	6 Pro	7 Gly	8 Hyp	9 Asp	10 Gly
AA (mg)	142.7	196.5	169.6	142.7	196.6	161.9	142.7	196.6	197.5	142.7
Oxyma (mg)	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0	70.0
DMF (mL)	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6
DIC (µL)	74.0	74.0	74.0	74.0	74.0	74.0	74.0	74.0	74.0	74.0

Table 5-3: Components of the coupling mixture for each aa used in the Fmoc-SPPS synthesis of the heavy T5 peptide.

F-moc AA residue	1 Gly	2 Hyp	3 $^{13}C/^{15}N$-Ile	4 Gly	5 Hyp	6 Pro	7 Gly	8 Hyp	9 Asp	10 Gly
AA (mg)	142.7	196.5		142.7	196.6	161.9	142.7	196.6	197.5	142.7
Oxyma (mg)	70.0	70.0	52.5	70.0	70.0	70.0	70.0	70.0	70.0	70.0
DMF (mL)	1.6	1.6	1.2	1.6	1.6	1.6	1.6	1.6	1.6	1.6
DIC (µL)	74.0	74.0	55.5	74.0	74.0	74.0	74.0	74.0	74.0	74.0

Table 5-4: Components of the coupling mixture for each aa used in the Fmoc-SPPS synthesis of the *h*T1 peptide.

F-moc AA residue	1 Asp	2 Ala	3 Tyr	4 Ser	5 Gln	6 Gly	7 Leu	8 His	9 Ser	10 Cys	11 Gly	12 Gly	13 Glu	14 Val	15 Ala	16 Glu	17 Gln	18 pyroE
AA (mg)	59.2	44.8	66.1	55.2	88.9	42.8	50.9	89.2	55.2	84.3	42.8	42.8	61.3	48.9	44.8	61.3	88.9	46.5
Oxyma (mg)	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	-
DMF (mL)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1.6
DIC (µL)	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	-
HATU (g)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	130.0
DIPEA (µL)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	125.4

5.3.1.2 LC-UV analysis

To monitor the presence of impurities produced in the peptide product, a sample of the crude product (~100 µg/mL, 5 µL) in HCOOH (0.3 %) and CH_3CN (5 %) solution was analysed by LC—UV using a Hewlett Packard (Agilent Technologies. Santa Clara, CA, USA) series 1050 HPLC coupled to a Hewlett Packard 1046A programmable fluorescence detector (Waldbronn, Germany). Gradient LC separation was achieved on an Agilent Technologies Zorbax SB-C-18 (3.5 µm, 2.1 x 100 mm) column with mobile phase A, 0.1 % trifluroacetic acid in water, and B, 0.1 % trifluroacetic acid in acetonitrile and a flow rate of 300 µL/min. At the start of the gradient mobile phase B was 0 % and increased to 90 % over 20 min, where it was held for a further 3 min. Mobile phase B was then decreased to 0 % over 2 min and held for 6 min to re-equilibrate the column. UV absorbance at 254, 230, 214, 220 and 281 nM were monitored (with 550 nM as a reference) for all signals.

5.3.1.3 LC-MS/HRMS

To ensure the desired peptides were within their synthesised product, full and the $[M+2H]^{2+}$ product ion scan spectra of each product (~1 µg/mL, 5 µL) in formic acid (0.3 %) and acetonitrile (5 %) solution were collected (where possible) by LC-MS/HRMS using the conditions detailed in section 2.3.3 and section 2.3.4. For the heavy T5 peptide the $[M+2H]^{2+}$ precursor ion corresponds to *m/z* 538.

5.3.1.4 Purification of synthesised peptide product

Fractionation of the crude peptide product in formic acid (0.3 %) and acetonitrile (5 %) solution (5mg/mL, 8mL, injected at a flow of 1 mL/min) was achieved on a Waters® HPLC system (515 HPLC pump equipped with a binary gradient module 2545, 267 sample manager and solvent flow operator) coupled to a Waters® Micromass ZQ and Waters® 2996 Photodiode array detector. Separation was achieved on a Waters® XTerra™ MS C18 (3.5 µm, 4.6 x 150mm) column with mobile phase A, 0.1 % formic acid, 5 % acetonitrile in water, and B, 0.1 % formic acid in acetonitrile and a flow rate of 10 mL/min. Gradient elution was achieved over 30 min, where the % B was changed

from 0 to 70 %, and then decreased to 10 % in 5 min, where it was held for 3 min before decreasing to 0 % in 2 min before equilibrating the column for a further 8 min.

For MS analysis the instrument was operated in positive ESI mode with a capillary and cone voltage were 2.7 kV and 30 V, respectively. The source and desolvation temperatures were 150 and 300 °C. The cone and desolvation gas flows were 50 and 500 L/H. Fractionation of sample was triggered at selected m/z (± 1 amu) whilst scanning over a range of m/z 300 to 1100, and UV at 214 nM with a diode array detecting wavelength between 210 to 500nM.

5.3.2 *Quantification of peptide by aa analysis*

In duplicate, the purified T5 peptide product (956 µg/mL, 500 µL) was hydrolysed in hydrochloric acid (6M) at 110 °C for 24 h using pressurised glass vessels. The hydrolysed sample was diluted 1 in 100 with acetonitrile containing 5 nM ^{13}C-Arg and ^{13}C-Leu. The sample was analysed by LC-MS using the targeted aa LC-MS method described below, and compared to aa standards for quantification. The experiment was repeated in a glass test tube for comparison.

5.3.2.1 Amino acid analysis

Separation of the amino acid standard mix (500 pM - 50 nM, 10 µL) and the hydrolysed peptide sample solution (both in acetonitrile containing 5 nM ^{13}C-Arg and ^{13}C-Leu) was achieved on a Waters Acquity® UPLC® coupled to a Xevo® TQ-S MS interfaced with a conventional ESI ion source using the chromatographic separation described by Hubertus *et al.* [189]. Gradient LC separation of the aa standard (10 µL) was achieved using an Acquity® UPLC® BEH Amide (1.7 µm, 2.1 x 150 mm) column at 35 °C with a flow of 400 µL/min. Mobile phase A (10 mM ammonium formate in 85 % acetonitrile containing 0.2 % formic acid) and B (10 mM ammonium formate in water containing 0.2 % formic acid) were used. Mobile phase A was held at 100 % for 6 min, after a which a gradient was started and mobile phase decreased to 94.1 % in 0.1 min, before decreasing to 82.4 % in 3.9 min and then to 70.6 % in 2 min. Mobile phase A was then returned to 100 % in 0.1 min, and the column re-equilibrated for 4 min.

For MS analysis the instrument was operated in positive mode with a capillary voltage of 2.8 kV. The source and desolvation temperatures were 150 and 200 °C. The cone flow 150 L/h and desolvation gas flow 0.15 mL/min. SRM transitions monitored for the respective aa residues are shown in Table 5-5 below.

Table 5-5: Initial cone voltages and collision energies for the SRM (selected reaction monitoring) transitions acquired for amino acid standards using a Waters Xevo®-TQS.

Amino acid	Precursor ion (*m/z*)	Product ion (*m/z*)	Cone Voltage (V)	Collision energy (eV)
Leu	132	44	32	16
	132	69	32	14
	132	86	32	8
Heavy Ile	138	91	32	12
Arg	175	60	8	12
	175	70	8	18
	175	116	8	12
Heavy Arg	181	74	8	20
Asp	134	74	32	10
	134	88	32	15
Pro	115	70	20	10

5.4 **Results and Discussion**

The preliminary results are presented for the peptide synthesis of the isotopic heavy-labelled ($^{13}C/^{15}N$) and unlabelled P-III-NP *h*T1 and T5 peptides analogues. To assess the applicability of the chosen resin for synthesis unlabelled peptide analogues were synthesised first. For T5, both analogues were synthesised and have been purified, however they await quantification by aa LC-MS analysis. For T1, however, only synthesis of the unlabelled analogue has been achieved so far. Though the aa quantification methodology is incomplete, the process of development of this method is presented and discussed.

5.4.1 ***Synthesised T5 peptide product***

5.4.1.1 Purity

Using the absorbance of peptide bonds at a wavelength of 214 nM, absorbance readings for the peptide product can be used to detect the number of components in a peptide mixture after LC separation, thus providing a purity profile. Using this approach, chromatograms representative of the purity of the crude and fractionated (purified) peptide product have been attained by LC-UV analysis and are presented in Figure 5-1 and Figure 5-2 below.

In the crude peptide product, several peptides are detected along with the peptide product (t_R 7.5 min, Figure 5-1). These additional products most likely originate from deleted, truncated or incompletely de-protected peptide sequences produced during synthesis. To remove and purify for the desired product, fractionation of the LC separated product was performed, for which a combined UV (214 nM) and MS (*m/z* 534 ± 1 amu) trigger were employed. Purity analysis of the purified peptide product, show one peak at t_R 7.4 min (Figure 5-2).

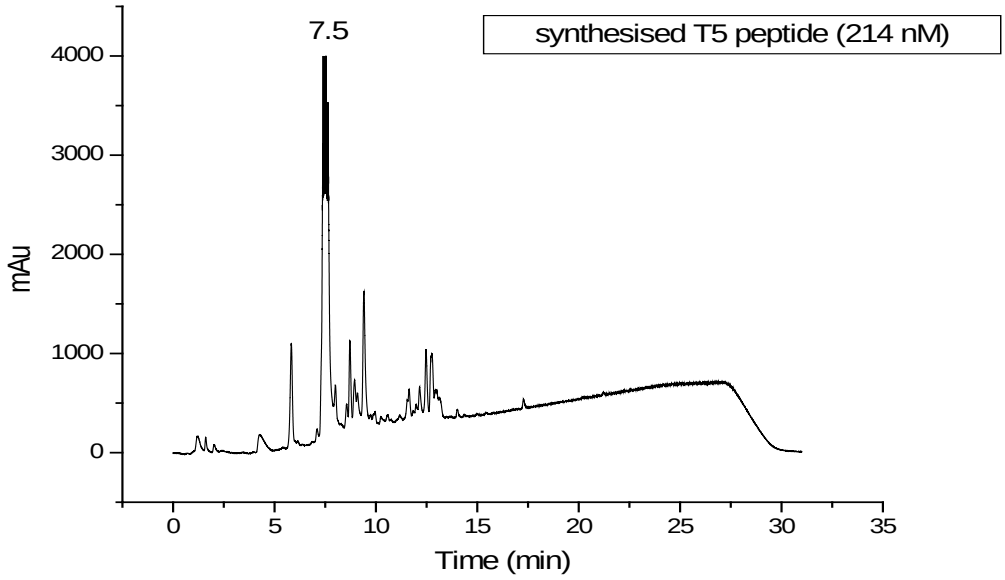


Figure 5.4-1: Chromatogram of the crude T5 peptide product obtained by LC-UV analysis monitoring absorbance at 214 nM.

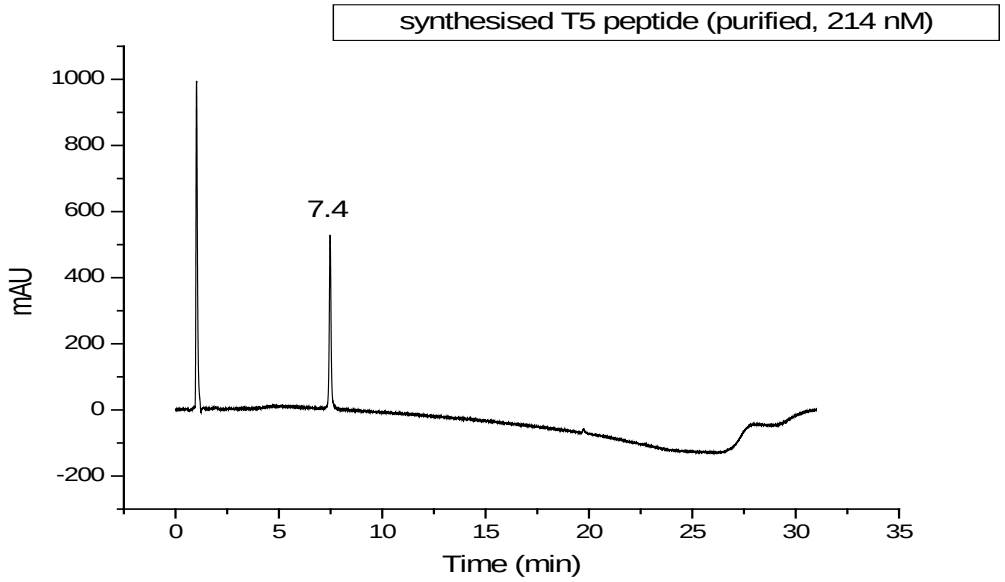


Figure 5.4-2: Chromatogram of the purified T5 peptide product obtained by LC-UV analysis monitoring absorbance at 214 nM.

5.4.1.2 Sequence verification

In SPPS, the Wang-linked resin was used for synthesis, hence the cleaved peptide product should have a carboxyl and amino group at the respective ends of the peptide.

Analysed by LC-MS/HRMS the crude and purified peptide products give the full and product ion scan spectra shown in Figure 5-3 and Figure 5-4. In the MS full scan, the singly- and doubly-charged T5 peptide ions were observed within the spectrum of the peak eluting at t_R 4.25 min, with a *m/z* 1067.5096 (theoretical *m/z* 1067.5116) and *m/z* 534.2596 (theoretical *m/z* 534.2594), respectively. The b- and y-ions observed in the product scan spectrum of the doubly-charged peptide ion (534.26 ±0.2 amu) correspond with the aa sequence $GDP_{+OH}GPP_{+OH}GIP_{+OH}GR$, which is the desired sequence for T5.

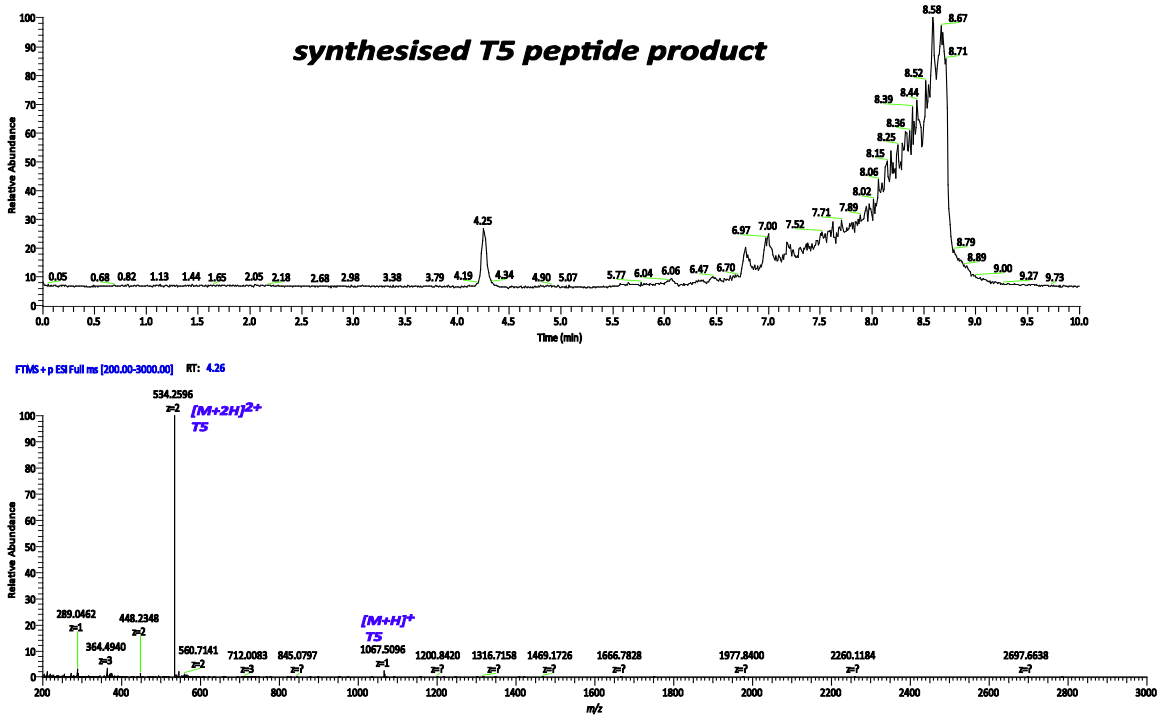


Figure 5.4-3: Total ion chromatogram of the synthesised T5 peptide product showing the full scan MS spectrum at t_R 4.25 min where the $[M+H]^+$ and $[M+2H]^{2+}$ T5 peptide ions are annotated.

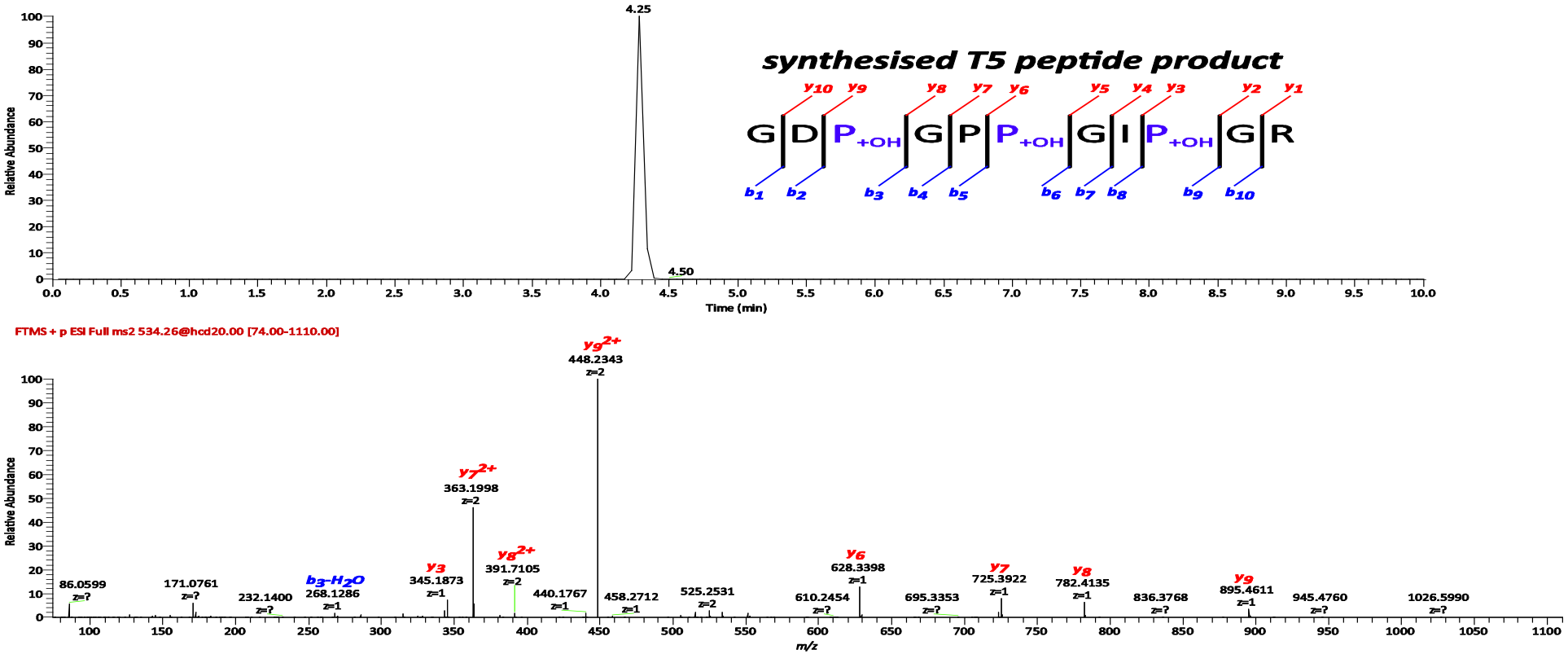


Figure 5.4-4: Extracted ion chromatogram at *m/z* 534.2594 ±5 ppm and corresponding product scan spectrum of the T5 doubly-charged peptide ion using a normalised collision energy of 20 %. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged. (Refer to Table 2-10 for theoretical *m/z* values).

5.4.2 *Synthesised heavy T5 peptide product*

5.4.2.1 Purity

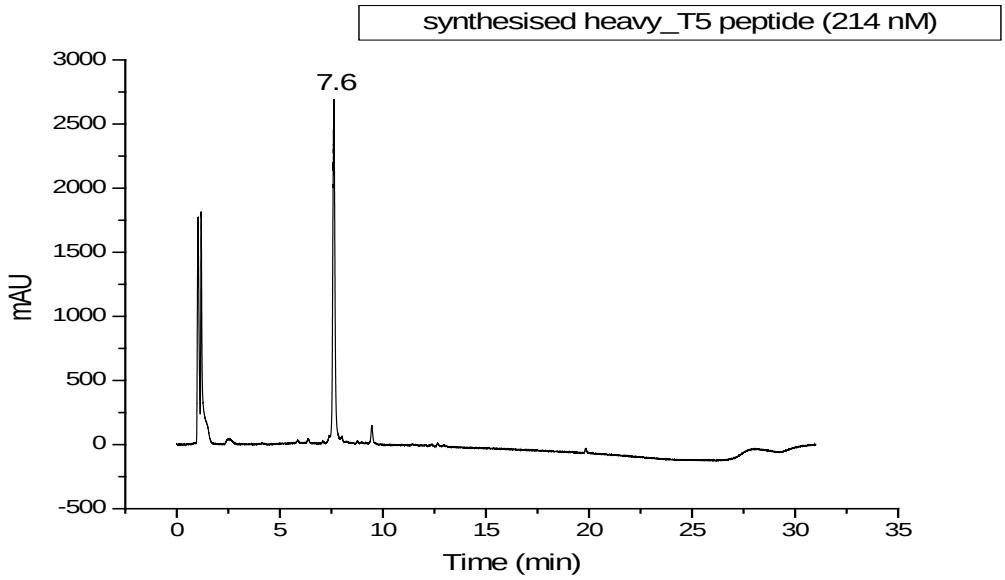


Figure 5.4-5: Chromatogram of the crude heavy T5 peptide product obtained by LC-UV analysis monitoring absorbance at 214 nM.

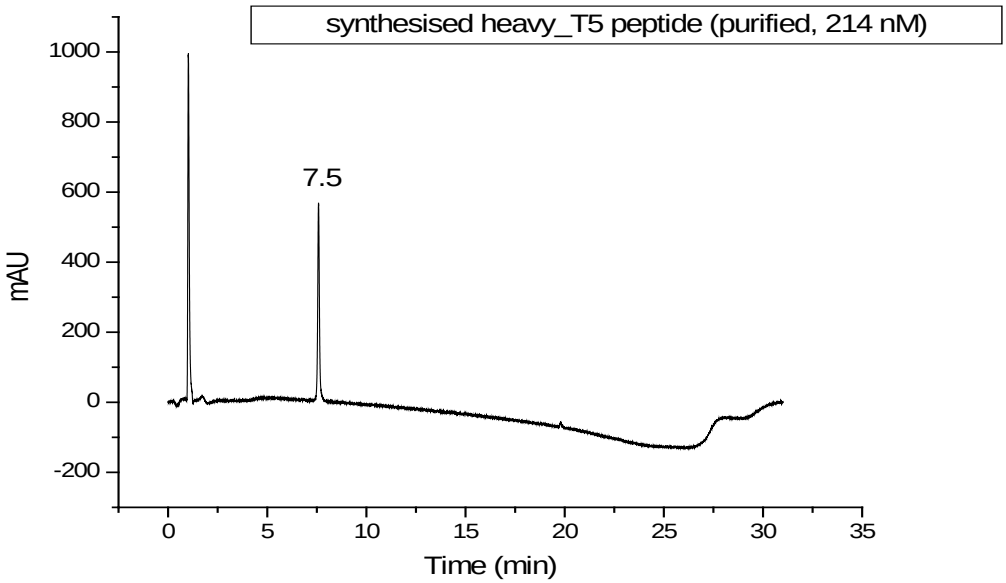


Figure 5.4-6: Chromatogram of the purified heavy T5 peptide product obtained by LC-UV analysis monitoring absorbance at 214 nM.

Chromatograms representative of the purity profile for the crude and purified heavy T5 products are presented in Figure 5-5 and Figure 5-6 above. Compared to the crude product of T5, a small amount of peptide impurity was observed in the heavy T5 product, nevertheless, fractionation of the sample was performed using LC with combined UV (214 nM) and MS (*m/z* 538 ± 1 amu) triggered separation. Re-analysed by LC-UV, a single peptide is observed at t_R 7.5 min (Figure 5-6).

5.4.2.2 Sequence verification

Analysed by LC-MS/HRMS the crude and purified peptide product give the full and product ion scan spectra shown in Figure 5-7 and Figure 5-8 below. In the MS full scan, the expected singly- and doubly-charged heavy T5 peptide ions were observed within the spectrum of the peak eluting at t_R 4.25 min, at *m/z* 1074.5275 (theoretical *m/z* 1074.5288) and *m/z* 537.7680 (theoretical *m/z* 537.7680), respectively. The b- and y-ions observed in the product scan spectrum of the doubly-charged peptide ion (534.26 ±0.2 amu) correspond with the aa sequence $GDP_{+OH}GPP_{+OH}G^{13}C_6/^{15}N\text{-}IP_{+OH}G$; which is the desired sequence for heavy T5.

Co-elution of the T5 (Figure 5-3) and heavy T5 (Figure 5-7) peptides was observed at t_R 4.25 min. This is expected as both peptides have similar aa sequence, hence they show comparable affinities for the C-18 chemistry of the LC column. Differentiation of the two peptides is achieved by MS where the presence of the isotopic $^{13}C_6/^{15}N$-Ile residue replaces the Ile residue in T5. This associates a +7 Da shift in the peptide mass of heavy T5 and is evidenced with the observation of the $[M+H]^+$ ion at *m/z* 1074.5274 for (compared to T5, $[M+H]^+$ *m/z* 1067.5116). Based on the location of the isotopic residue in the heavy T5 peptide, conservation of the appropriate mass shift (+7 and +3.5 Da for the singly- and doubly-charged peptide ions) is expected for all the SRM transitions monitored in the nano-and micro-flow LC-MS targeted P-III-NP methods (Section 3.3.1.1, Table 3-3). For heavy T5, the y_7^{2+}, y_9^{2+} and y_6^{+} fragment peptide ions are observed at *m/z* 366.7089, 451.7435 and 635.3576 (compared to T5 peptide ions *m/z* 363.2007, 448.2352 and 628.3413), respectively.

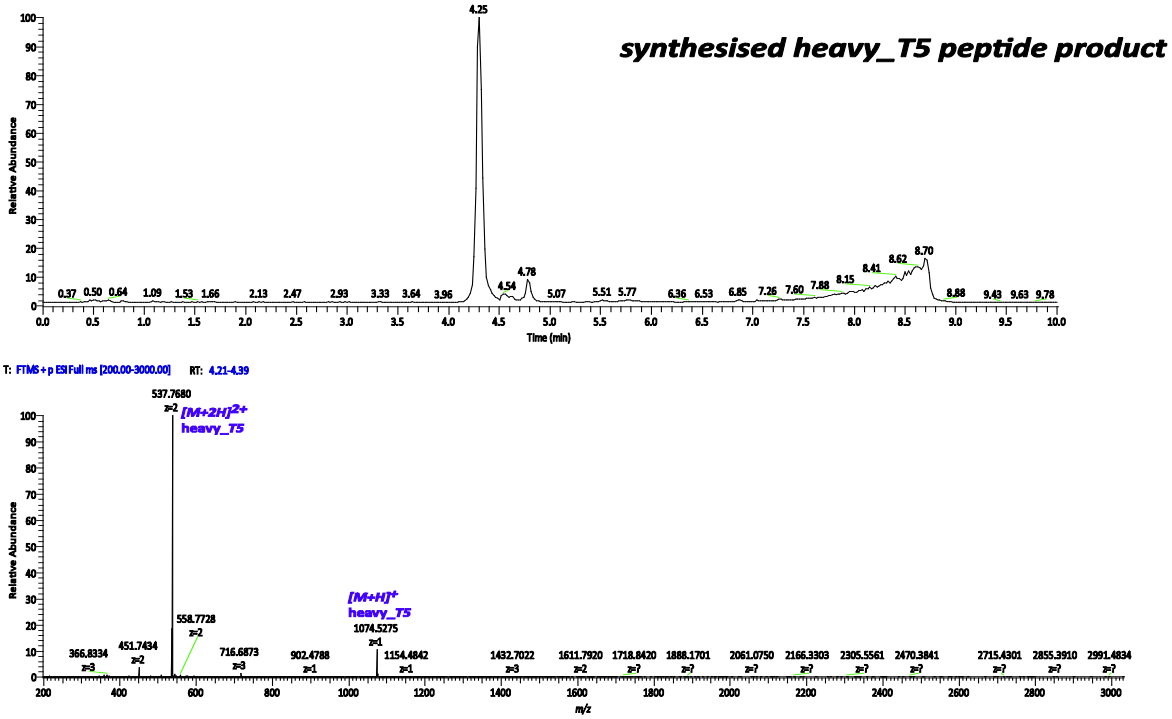


Figure 5.4-7: Total ion chromatogram of the synthesised heavy T5 peptide product showing the full scan MS spectrum at t_R 4.25 min where the $[M+H]^+$ and $[M+2H]^{2+}$ heavy T5 peptide ions are annotated.

Table 5-6: Theoretical *m/z* of the b- and y-ions of heavy T5 peptide ion.

b- fragment				y- fragment		
m/z **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**	**fragment no.**	**amino acid sequence**	**fragment no.**	***m/z*** **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**
-	-	1	**G**	11	-	-
173.0557	-	2	**D**	10	1017.5073	509.2573
286.1034	-	3	**P_{+OH}**	9	902.4804	451.7438
343.1248	-	4	**G**	8	789.4327	395.2200
440.1776	-	5	**P**	7	732.4112	366.7093
553.2253	-	6	**P_{+OH}**	6	635.3585	318.1829
610.2467	-	7	**G**	5	522.3108	261.6590
730.3480	-	8	**$^{13}C_6/^{15}N$-I**	4	465.2893	233.1483
843.3956	-	9	**P_{+OH}**	3	345.1881	173.0977
900.4171	-	10	**G**	2	232.1404	116.5738
-	-	11	**R**	1	175.1190	88.0631

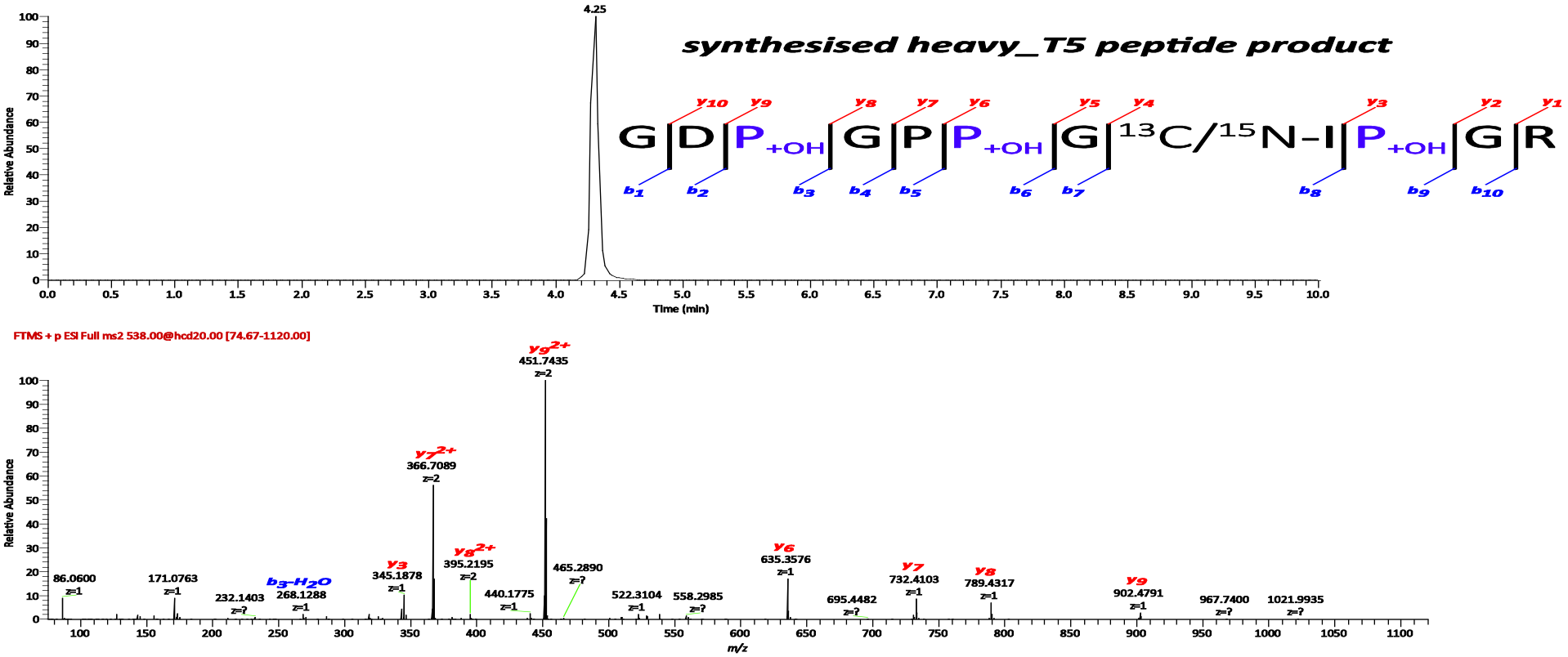


Figure 5.4-8: Extracted ion chromatogram at *m/z* 537.7680 ±5 ppm and corresponding product scan spectrum of the heavy T5 doubly-charged peptide ion using a normalised collision energy of 20 %. The most abundant b- and y-ions are annotated, where no charged is indicated the ions are singly-charged. (Refer to Table 5-6for theoretical *m/z* values).

5.4.3 ***hT1 peptide***

5.4.3.1 Purity

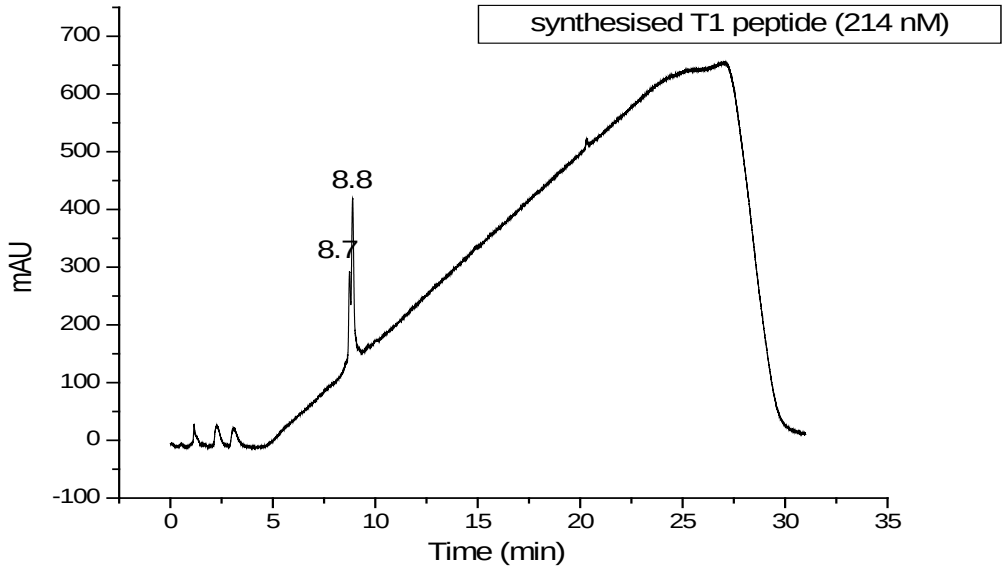


Figure 5.4-9: Chromatogram of the crude *h*T1 peptide product obtained by LC-UV analysis monitoring absorbance at 214 nM.

The chromatogram of the crude synthesised *h*T1 peptide product obtained by LC-UV is shown in Figure 5-9, where two eluting peaks with low resolution are observed at t_R 8.7 and 8.8 min. Unlike the T5 peptides, *h*T1 contains a Cys residue that is readily oxidised to form a disulphide bond to result in the formation of a peptide dimer. Thus, these closely eluting peaks may originate from the same peptide in different oxidative states, as opposed to being present as a result of being a synthesis impurity.

In the purification of this peptide, differences in *m/z* of the reduced and oxidised peptide forms need to be considered so as to maximise on the recovery of the peptide product. Complete reduction and alkylation of the *hT1* peptide will simplify the purification of the peptide with the production of a single MS target (corresponding to the derivatised peptide), however, the stability of the alkylated peptide is unknown. Hence, purification of the underivatised peptide product is recommended. *h*T1 peptide synthesis is used as a model for the heavy variant, hence once optimisation of purification methods are complete, synthesis of the heavy *h*T1 will commence.

5.4.3.2 Sequence verification

In the full scan spectrum (Figure 5-10) of the crude synthesised product only the reduced *h*T1 peptide was observed. The $[M+2H]^{2+}$ ion at *m/z* 1009.4353 (theoretical *m/z* 1009.4373) is present in the spectrum although dominated by a more abundant isotope at *m/z* 1009.9367. Also the $[M+3H]^{3+}$ ion at *m/z* 673.2925 (theoretical *m/z* 673.2941) is observed in the spectrum. The absence of the oxidised peptide could be due to low concentration of injected sample, which has been diluted 100-fold compared to that analysed by LC-UV. Thus the analysis of a more concentrated sample is to be undertaken. Using the reduced doubly-charged peptide ion as precursor, the product ion spectrum (Figure 5-11) obtained using an nCE of 20 % although showing a dominant peak for the precursor ion, the b- and y-peptide ions shown corresponds with that of the desired *h*T1 aa sequence Q_{-NH3}QEAVEGGCSHLGQSYADR. It is expected that if the collision energy is increased, the intensity of the observed peptide fragments ions will also increase, however this still to be confirmed.

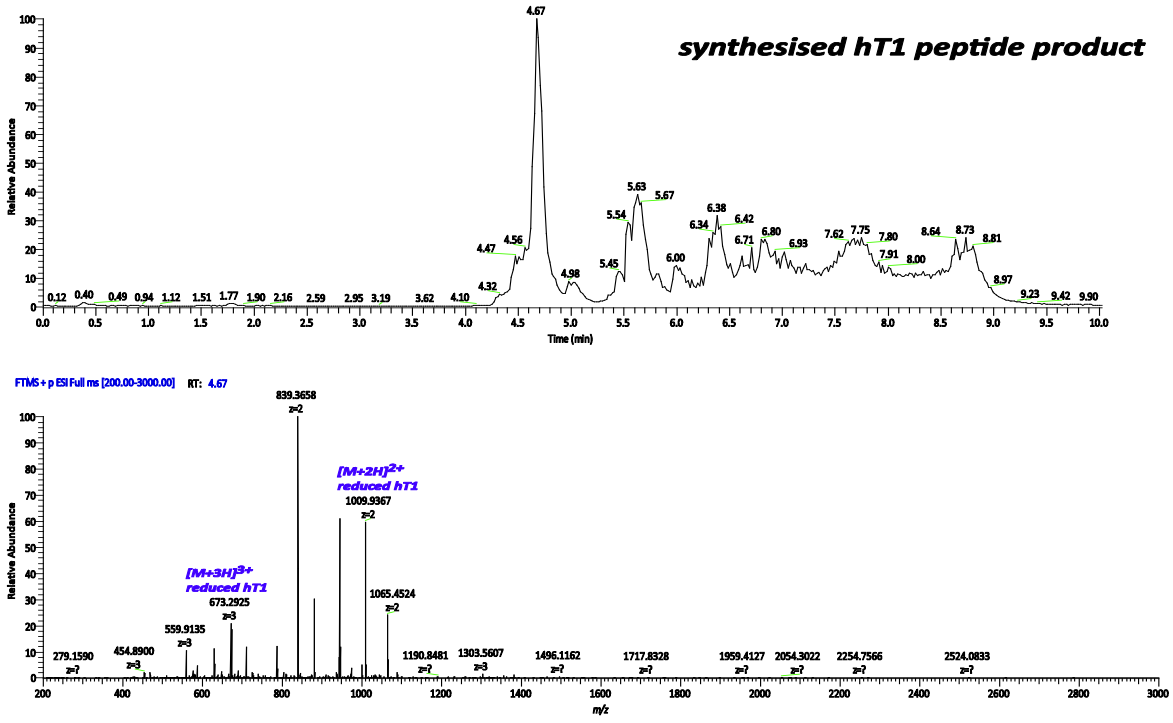


Figure 5.4-10: Total ion chromatogram of the crude synthesised *h*T1 peptide product showing the full scan MS spectrum at t_R 4.68 min where the $[M+2H]^{2+}$ and $[M+3H]^{3+}$ hT1 peptide ions are annotated

Table 5-7: Calculated *m/z* of the b- and y-ions of reduced *h*T1 peptide ion.

b- fragment				y- fragment		
m/z **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**	**fragment no.**	**amino acid sequence**	**fragment no.**	***m/z*** **$[M+H]^+$**	***m/z*** **$[M+2H]^{2+}$**
-	-	1	**Q_{-NH3}**	19	-	-
240.0979	-	2	**Q**	18	1906.8348	953.9210
369.1405	-	3	**E**	17	1778.7762	889.8918
440.1776	-	4	**A**	16	1649.7336	825.3705
539.2460	-	5	**V**	15	1578.6965	789.8519
654.2729	-	6	**D**	14	1479.6281	740.3177
711.2944	-	7	**G**	13	1350.5855	675.7964
768.3159	-	8	**G**	12	1293.5641	647.2857
884.3407	-	9	**C**	11	1236.5426	618.7749
972.3727	**-**	10	**S**	10	1133.5334	567.2703
1109.4316	555.2195	11	**H**	9	1046.5014	523.7543
1221.5317	611.7615	12	**L**	8	909.4425	455.2249
1279.5372	640.2722	13	**G**	7	796.3584	398.6828
1407.5958	704.3015	14	**Q**	6	739.3369	370.1721
1561.6700	747.8175	15	**S**	5	611.2784	306.1428
1632.7071	829.3492	16	**Y**	4	524.2463	262.6268
1728.7282	864.8678	17	**A**	3	361.1830	181.0951
1843.7552	922.3812	18	**D**	2	290.1459	145.5766
-	-	19	**R**	1	175.1190	88.0631

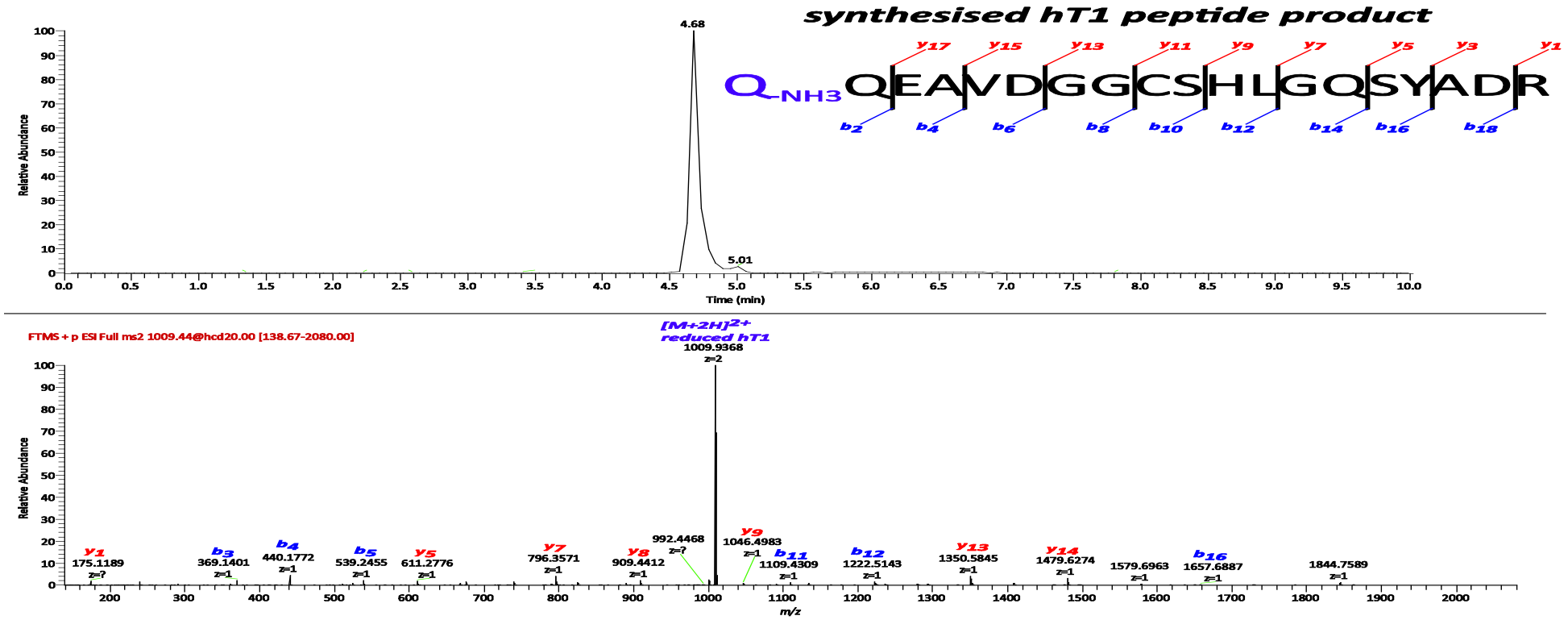


Figure 5.4-11: Extracted ion chromatogram at *m/z* 1009.4373 ±5 ppm and corresponding product scan spectrum of the reduced *h*T1 doubly-charged peptide ion using a normalised collision energy of 20 %. Some b- and y-ions are annotated, where no charged is indicated the ions are singly-charged. (Refer to Table 5-7 for theoretical *m/z* values).

5.4.4 ***Protein quantification by aa analysis***

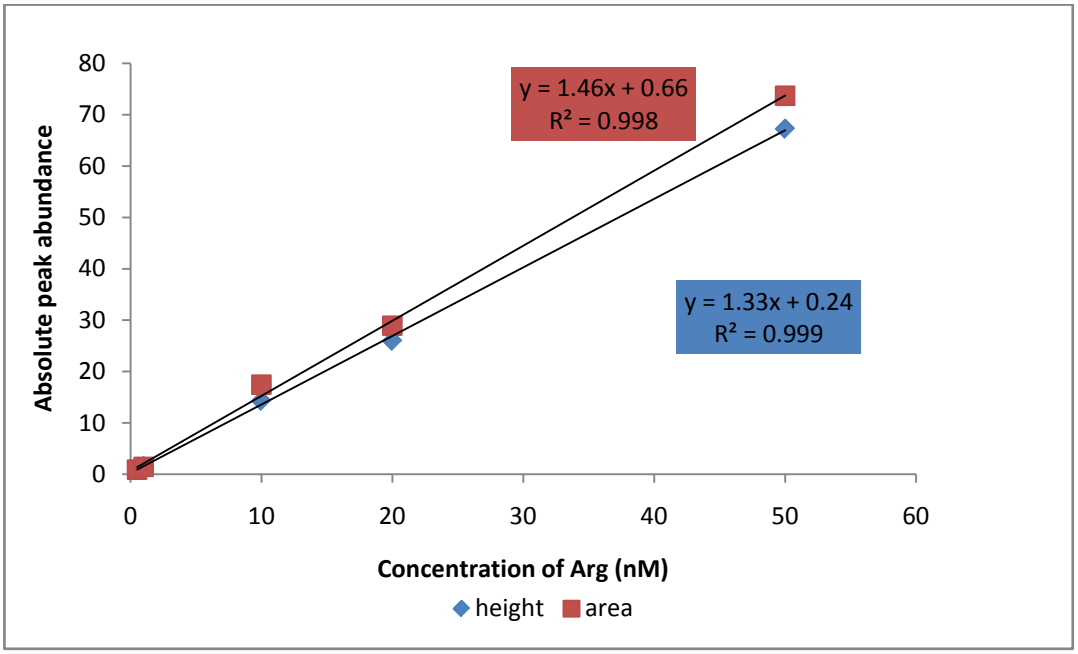


Figure 5.4-12: Standard curve of Arg (nM) using absolute peak area and height obtained by LC-MS analysis. The equation and regression of each curve is shown.

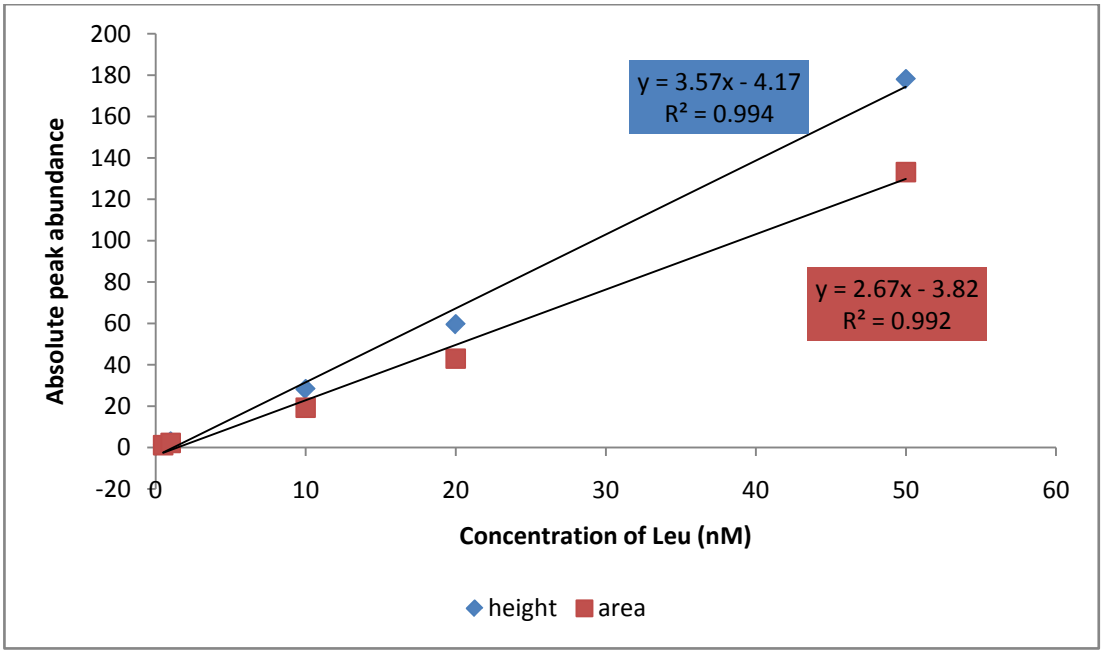


Figure 5.4-13: Standard curve of Leu (nM) using absolute peak area and height values by LC-MS analysis. The equation and regression of each curve is shown.

Using the standard curves for Arg and Leu (shown in Figure 5-12 and Figure 5-13, respectively) the concentration of the acid hydrolysed purified synthesised T5 product (theoretical concentration 4.48 nM) was determined. The concentration of each sample is shown in Table 5-8 below. Of the aa chosen for quantification, Ile seems to be more precise, and produces an average peptide concentration of 4.24 nM (using peak height). For Arg, less precision is observed within the replicates, and a lower concentration than expected was obtained from the samples hydrolysed in the pressurised glass container, as opposed to the test tube. This could be due to loss of analyte to the surface of the vessel, or incomplete hydrolysis of the sample. The concentration of Arg obtained from the peptide prepared in the test tube is 4.66 nM (using peak area). This vessel is thinner than the pressurised glass, hence the efficiency of heat conduction could be greater. Based on the location of each residue in the sequence, it would be expected that Arg is cleaved first. As a free aa in the reaction mixture, adsorption losses are a more likely cause of the reported low Arg concentration. Hence investigation into the degree of loss and methods to correct or avoid this are needed. In addition the use of other aa (e.g. Gly) for quantification should be investigated.

Table 5-8: Absolute concentration values of synthesised T5 peptides using aa analysis. Where (PG) and (G) represents pressurised glass or test tube, respectively, to indicate the vessel in which the sample was hydrolysed.

	Arg Concentration (nM)		**Ile concentration (nM)**	
	Area	*Height*	*Area*	*Height*
Sample 1 (PG)	3.49	3.32	4.78	4.10
Sample 2 (PG)	2.08	2.03	5.35	4.33
Sample 3 (G)	4.66	7.10	4.94	4.31

5.5 **Summary**

Peptides for T5, heavy T5 and *h*T1 have been produced by solid phase protein synthesis. The *de novo* peptide sequence obtained from the LC-MS/HRMS analysis of each product verifies that the aa sequence is that of the desired peptide. Purification of

the T5 and heavy T5 peptides was successful and methods are being developed for quantification. Possessing a more complex chemistry due to its vulnerability to oxidation, purification methods for the *h*T1 peptide still requires optimisation. Once completed, synthesis of the heavy_*h*T1 peptide will commence.

Chapter 6 *: Conclusions and future work*

6.1 **Conclusions**

As an initial step towards developing a quantitative method based on mass spectrometry for analysing P-III-NP, the amino acid sequence of the protein needs to be fully characterised. Proteins often incorporate structural elements associated with the post-translational modifications (PTMs) of some of the genetically encoded aa in their sequence. These PTMs are challenging to replicate in synthesis (e.g. specific Cys-Cys disulphide bridges or glycosylated units), hence, even with knowledge of the aa sequence synthetic protein synthesis is impossible. Recombinant DNA technologies may be considered for synthesis where an appropriate cell line has been selected to support PTMs and correct topological formation of the expressed protein.

Based on the size of the P-III-NP, full characterisation is difficult by MS. However with the removal of the protein's disulphide bonds (especially the intra-chain bonds at the carboxyl-terminal end) to yield three identical pro α-1 chains that can be further digested with trypsin to yield reproducible peptides suitable for MS analysis partial protein characterisation has been possible. By this simplification, peptide sequences with PTMs are produced that lack the structural characteristics of the intact protein, which are less complicated to synthesise. This enables the production of reference material for unique P-III-NP peptides. Both as normal and heavy labelled variants, in quantities that allow for certification as quantitative standards usually required for doping control purposes.

Identification of P-III-NP peptide fragments T1 and T5 by HRMS was obtained for the first time and is presented here, helping advance the characterisation of P-III-NP. Based on the structure of P-III-NP, some PTMs can be predicted. The data presented in Chapter 2 highlights the importance of including PTMs in protein MS identification, as sole dependence on generated *in silico* digest fragments can lead to the non-detection of peptide fragments.

Once selected, specific P-III-NP peptides *h*T1 and T5, were used to develop sensitive micro- and nano-flow LC-MS analytical methods, for which an LOD as low as 5 pM (5 µL) was achieved and is presented in Chapter 3. However, without extraction from serum, the analysis of P-III-NP by LC-MS was unachievable, as a result of ion suppression from the abundant albumin concentration in the matrix. Thus the applicability of use of some common proteomic sample preparation approaches (e.g. protein precipitation, filtration and immunocapture) for separation of P-III-NP and albumin were investigated as described in Chapter 4. Here successful immunocapture of P-III-NP from a pooled normal human serum sample enabled the identification of *h*T1 and T5 peptides by LC-MS analysis after tryptic digestion, for the first time.

To determine P-III-NP concentrations from LC-MS measurement, parallel quantification of *h*T1 and T5 peptides is desired. Using this comparative evaluation of the peptide concentrations increases the accuracy of P-III-NP measurements. Unfortunately, further optimisation of the assay selectivity for *h*T1 is needed before *h*T1 measurements can be considered for quantification. In the absence of this, T5 semi-quantification of the analysed serum sample gave a protein concentration of 2 ng/mL (50 pM), which is within the expected basal serum range of P-III-NP (1-5 ng/mL, 25-125 pM).

To develop commutable quantitative LC-MS methods, standardised international reference material for P-III-NP (as peptides or intact protein) as well as heavy-labelled “internal standards” are needed. In chapter 5 the initial steps that have been taken towards the "in-house" synthesis of labelled and unlabelled *h*T1 and T5 peptides are described. Here, high standards for purification and qualification of these standards have been adopted to ensure reliability in their intended use for quantification. With the completed synthesis of these peptide standards, full quantification of P-III-NP by LC-MS will be possible, which will allow for the comparison of measurements with those determined by immunoassay approaches.

6.2 **Future work**

The work presented within this thesis clearly demonstrates the promises of LC-MS for P-III-NP analysis. To deliver accurate measurements, the comparative quantification of unique tryptic peptides *h*T1 and T5 is intended, however further optimisation of *h*T1 analysis to improve its separation from HSA and investigate the completion of derivatisation to increase sensitivity, is needed before this approach can be used. For protein quantification, measurement of two peptides is not necessary hence in the absence of assay specificity for *h*T1, T5 based measurements should be assessed. For quantitative peptide analysis, however, the completed synthesis and qualification of reference standards is imperative to assay development. Hence, as the next step towards quantification, continuation of the peptide synthesis already initiated (see Chapter 5) is necessary.

Once quantification of serum P-III-NP is achieved by the developed LC-MS approach, comparison of assay results with those determined by immunoassay should be used to validate assay measurements. Within this study a comprehensive statistical analysis of LC-MS based measurements should be conducted to facilitate the determination of assay performance. Once a viable assay has been produced, the re-calculation of GH biomarker scores based on the LC-MS measurements of P-III-NP should be considered. This will for the first time produce scores independent of immunoassay measurements and allow for full comparison of the discriminatory factor of the two approaches.

As a digest approach for quantification has been adopted, the use of more readily synthesised peptide standards is applicable in the absence of authentic intact protein. To improve the sensitivity of MS measurements, incorporation of intact protein standard in the sample flow will normalise sample preparation losses to improve assay results. Hence, as a way forward, the production of an intact P-III-NP standard, as labelled and unlabelled protein, is valuable. To achieve this by SPPS, full characterisation of the protein is necessary so that all PTMs may be considered. In

addition to the incorporation of PTMs, unlike the peptide synthesis, the correct cross-linking of the pro-α1 chains is necessary post SPPS to complete synthesis. This is a challenging and complicated step, which will produce several isomers and affect the purity of the protein product to result in a low production yield. As discussed in Chapter 1, recombinant procollagen III can be produced where r*h*P-III-NP is a waste by-product of the target collagen III. Unfortunately current synthesis excludes the P-III-NP unit from the recombinant product to increase the yield of the target protein. If intact P-III-NP becomes available, it is recommended that the peptide sequences of *h*T1 and T5 are checked using the developed methods, to ensure compatible sequence with the endogenous P-III-NP protein.

Similar to immunoassays, the immunocapture method currently used for sample preparation, is vulnerable to antibody modification or withdrawal by manufacturers. To overcome this threat and enable method viability, alternative sample preparation techniques should be developed. Applicability of peptide or protein specific aptamers (single stranded nucleic acid) should be assessed for this purpose, along with other albumin depletion methods.

Chapter 7 *: References*

1. *The World Anti-Doping Agency International Standard Code: Prohibited List* T.W.A.-d. Agency, Editor. 2018, The World Anti-doping Agency.
2. Holt, R.I.G., I. Erotokritou-Mulligan, and P.H. Sönksen, *The history of doping and growth hormone abuse in sport.* Growth Hormone & Igf Research, 2009. **19**(4): p. 320-326.
3. Baumann, G.P., *Growth Hormone Doping in Sports: A Critical Review of Use and Detection Strategies.* Endocrine Reviews, 2012. **33**(2): p. 155-186.
4. Kicman, A.T., et al., *Serum IGF-I and IGF binding proteins 2 and 3 as potential markers of doping with human GH.* Clinical Endocrinology, 1997. **47**(1): p. 43-50.
5. Wallace, J.D., et al., *Responses of the Growth Hormone (GH) and Insulin-Like Growth Factor Axis to Exercise, GH Administration, and GH Withdrawal in Trained Adult Males: A Potential Test for GH Abuse in Sport*.* The Journal of Clinical Endocrinology & Metabolism, 1999. **84**(10): p. 3591-3601.
6. Beck, J.C., et al., *Metabolic effects of human and monkey growth hormone in man.* Science, 1957. **125**(3253): p. 884-5.
7. Leung, K.-C., et al., *Physiological and pharmacological regulation of 20-kDa growth hormone.* American Journal of Physiology - Endocrinology And Metabolism, 2002. **283**(4): p. E836-E843.
8. Baumann, G., *Growth hormone heterogeneity in human pituitary and plasma.* Hormone Research, 1999. **51**: p. 2-6.
9. Lecomte, C.M., A. Renard, and J.A. Martial, *A new natural hGH variant—17.5 kd—produced by alternative splicing. An additional consensus sequence which might play a role in branchpoint selection.* Nucleic Acids Research, 1987. **15**(16): p. 6331-6348.
10. Lewis, U.J., Y.N. Sinha, and G.P. Lewis, *Structure and Properties of Members of the hGH Family: A Review.* Endocrine Journal, 2000. **47**(SupplMarch): p. S1-S8.
11. Alsat, E., et al., *Human placental growth hormone.* American Journal of Obstetrics and Gynecology, 1997. **177**(6): p. 1526-1534.
12. Shuto, Y., et al., *Hypothalamic growth hormone secretagogue receptor regulates growth hormone secretion, feeding, and adiposity.* The Journal of Clinical Investigation, 2002. **109**(11): p. 1429-1436.
13. Muller, E.E., V. Locatelli, and D. Cocchi, *Neuroendocrine control of growth hormone secretion.* Physiological Reviews, 1999. **79**(2): p. 511-607.
14. Giustina, A. and J.D. Veldhuis, *Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human.* Endocrine Reviews, 1998. **19**(6): p. 717-797.
15. Hartman, M.L., *Physiological regulators of growth hormone.* Growth Hormone in Adults: Physiological and Clinical Aspects, 2000: p. 3.
16. Kojima, M., et al., *Ghrelin is a growth-hormone-releasing acylated peptide from stomach.* Nature, 1999. **402**(6762): p. 656-60.
17. Weltman, A., *Growth Hormone*, in *Hormone Use and Abuse by Athletes*, E. Ghigo, F. Lanfranco, and C.J. Strasburger, Editors. 2011, Springer US: Boston, MA. p. 89-98.
18. Winer, L.M., M.A. Shaw, and G. Baumann, *Basal plasma growth hormone levels in man: new evidence for rhythmicity of growth hormone secretion.* J Clin Endocrinol Metab, 1990. **70**(6): p. 1678-86.
19. Hartman, M.L., et al., *Temporal structure of in vivo growth hormone secretory events in humans.* Am J Physiol, 1991. **260**(1 Pt 1): p. E101-10.
20. Gibney, J., M.-L. Healy, and P.H. Sönksen, *The Growth Hormone/Insulin-Like Growth Factor-I Axis in Exercise and Sport.* Endocrine Reviews, 2007. **28**(6): p. 603-624.

21. Holt, R.I.G. and P.H. Sönksen, *Growth hormone, IGF-I and insulin and their abuse in sport.* British Journal of Pharmacology, 2008. **154**(3): p. 542-556.
22. Hansen, T.K., *Pharmacokinetics and acute lipolytic actions of growth hormone: Impact of age, body composition, binding proteins, and other hormones.* Growth Hormone & IGF Research, 2002. **12**(5): p. 342-358.
23. Li, C.H. and H. Papkoff, *Preparation and Properties of Growth Hormone from Human and Monkey Pituitary Glands.* Science, 1956. **124**(3235): p. 1293-1294.
24. Raben, M.S., *Treatment of a pituitary dwarf with human growth hormone.* The Journal of Clinical Endocrinology & Metabolism, 1958. **18**(8): p. 901-903.
25. Rennie, M.J., *Claims for the anabolic effects of growth hormone: a case of the Emperor's new clothes?* British Journal of Sports Medicine, 2003. **37**(2): p. 100-105.
26. Duchaine, D., *Underground Steroid Handbook*. 1982: OEM, Santa Monica, CA
27. Duchaine, D., *Ultimate Muscle Mass*. 1993: Mile High Publishing.
28. Giannoulis, M.G., et al., *The Effects of Growth Hormone and/or Testosterone in Healthy Elderly Men: A Randomized Controlled Trial.* The Journal of Clinical Endocrinology & Metabolism, 2006. **91**(2): p. 477-484.
29. Graham, M.R., et al., *Physical Effects of Short-Term Recombinant Human Growth Hormone Administration in Abstinent Steroid Dependency.* Hormone Research in Paediatrics, 2008. **69**(6): p. 343-354.
30. Goeddel, D.V., et al., *Direct expression in Escherichia coli of a DNA sequence coding for human growth hormone.* Nature, 1979. **281**(5732): p. 544-8.
31. Martial, J.A., et al., *Human growth hormone: complementary DNA cloning and expression in bacteria.* Science, 1979. **205**(4406): p. 602-7.
32. Weissberger, A.J., K.Y. Ho, and M.C. Stuart, *Quantification of urinary growth hormone (gh) excretion by centrifugal ultrafiltration and radioimmunoassay: Appraisal of the relationship between 24 h urinary gh and mean 24 h serum gh levels in normal and abnormal states of gh secretion.* Clinical Endocrinology, 1989. **30**(6): p. 687-698.
33. Wu, Z., et al., *Detection of doping with human growth hormone.* Lancet, 1999. **353**(9156): p. 895-895.
34. Irie, M., et al., *20K-GH and its use in detecting GH abuse.* Growth Hormone & IGF Research, 2009. **19**(4): p. 352-356.
35. Wallace, J.D., et al., *Changes in non-22-kilodalton (kDa) isoforms of growth hormone (GH) after administration of 22-kDa recombinant human GH in trained adult males.* Journal of Clinical Endocrinology & Metabolism, 2001. **86**(4): p. 1731-1737.
36. Giannoulis, M.G., et al., *Gender differences in growth hormone response to exercise before and after rhGH administration and the effect of rhGH on the hormone profile of fit normal adults.* Clinical Endocrinology, 2005. **62**(3): p. 315-322.
37. Dall, R., et al., *The effect of four weeks of supraphysiological growth hormone administration on the insulin-like growth factor axis in women and men.* Journal of Clinical Endocrinology & Metabolism, 2000. **85**(11): p. 4193-4200.
38. Longobardi, S., et al., *Growth hormone (GH) effects on bone and collagen turnover in healthy adults and its potential as a marker of GH abuse in sports: A double blind, placebo-controlled study.* Journal of Clinical Endocrinology & Metabolism, 2000. **85**(4): p. 1505-1512.
39. Kniess, A., et al., *Potential parameters for the detection of hGH doping.* Analytical and Bioanalytical Chemistry, 2003. **376**(5): p. 696-700.
40. Tapanainen, P., et al., *Serum aminoterminal propeptide of type-III procollagen - a potential predictor of the response to growth-hormone therapy.* Journal of Clinical Endocrinology & Metabolism, 1988. **67**(6): p. 1244-1249.
41. McHugh, C.M., et al., *Challenges in detecting the abuse of growth hormone in sport.* Clinical Chemistry, 2005. **51**(9): p. 1587-1593.
42. Powrie, J.K., et al., *Detection of growth hormone abuse in sport.* Growth Hormone & Igf Research, 2007. **17**(3): p. 220-226.

43. Healy, M.L., et al., *Toward the development of a test for growth hormone (GH) abuse: A study of extreme physiological ranges of GH-dependent markers in 813 elite athletes in the postcompetition setting.* Journal of Clinical Endocrinology & Metabolism, 2005. **90**(2): p. 641-649.
44. Nelson, A.E. and K.K.Y. Ho, *Demographic factors influencing the GH system: Implications for the detection of GH doping in sport.* Growth Hormone & IGF Research, 2009. **19**(4): p. 327-332.
45. Sönksen, P., *The International Olympic Committee (IOC) and GH-2000.* Growth Hormone & Igf Research, 2009. **19**(4): p. 341-345.
46. Erotokritou-Mulligan, I., et al., *The development of decision limits for the implementation of the GH-2000 detection methodology using current commercial insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays.* Growth Hormone & Igf Research, 2012. **22**(2): p. 53-58.
47. *Guidelines for Human Growth Hormone (hGH) Biomarkers Test for doping control analyses.* 2016, The World Anti-Doping Agency.
48. Holt, R.I.G., et al., *The development of decision limits for the GH-2000 detection methodology using additional insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays.* Drug Testing and Analysis, 2015: p. n/a-n/a.
49. Nelson, A.E. and K.K. Ho, *A robust test for growth hormone doping - present status and future prospects.* Asian Journal of Andrology, 2008. **10**(3): p. 416-425.
50. Rinderknecht, E. and R.E. Humbel, *The amino acid sequence of human insulin-like growth factor I and its structural homology with proinsulin.* J Biol Chem, 1978. **253**(8): p. 2769-76.
51. Clemmons, D.R., *IGF-I assays: current assay methodologies and their limitations.* Pituitary, 2007. **10**(2): p. 121-128.
52. Cowan, D.A. and C. Bartlett, *Laboratory issues in the implementation of the marker method.* Growth Hormone & IGF Research, 2009. **19**(4): p. 357-360.
53. Clemmons, D.R., *Clinical laboratory indices in the treatment of acromegaly.* Clinica Chimica Acta, 2011. **412**(5-6): p. 403-409.
54. Bystrom, C.E., S.J. Sheng, and N.J. Clarke, *Narrow Mass Extraction of Time-of-Flight Data for Quantitative Analysis of Proteins: Determination of Insulin-Like Growth Factor-1.* Analytical Chemistry, 2011. **83**(23): p. 9005-9010.
55. Bystrom, C., et al., *Clinical Utility of Insulin-Like Growth Factor 1 and 2; Determination by High Resolution Mass Spectrometry.* Plos One, 2012. **7**(9).
56. Lopes, F., et al., *Quantification of intact human insulin-like growth factor-I in serum by nano-ultrahigh-performance liquid chromatography/tandem mass spectrometry.* Rapid Communications in Mass Spectrometry, 2014. **28**(13): p. 1426-1432.
57. Cox, H.D., et al., *Interlaboratory Agreement of Insulin-like Growth Factor 1 Concentrations Measured by Mass Spectrometry.* Clinical Chemistry, 2014. **60**(3): p. 541-548.
58. Bentsen, K.D., *Type-III procollagen peptide - studies on the circulating peptide as a marker of fibrinogenesis with special reference to the liver.* Danish Medical Bulletin, 1993. **40**(2): p. 235-246.
59. Jensen, L.T., *The aminoterminal propeptide of type III procollagen - Studies on physiology and pathophysiology.* Danish Medical Bulletin, 1997. **44**(1): p. 70-78.
60. Kuhn, K., *The collagen family - variation of molecular and macromolecular structures.* Biological Chemistry Hoppe-Seyler, 1986. **367**(2): p. 88-89.
61. Hörlein, D., et al., *Regulation of protein synthesis: translational control by procollagen-derived fragments.* Proceedings of the National Academy of Sciences of the United States of America, 1981. **78**(10): p. 6163-6167.
62. Wiestner, M., et al., *Inhibiting effect of procollagen peptides on collagen biosynthesis in fibroblast cultures.* J Biol Chem, 1979. **254**(15): p. 7016-23.
63. Miller, E.J., et al., *Cleavage of Type II and III collagens with mammalian collagenase: site of cleavage and primary structure at the NH2-terminal portion*

of the smaller fragment released from both collagens. Biochemistry, 1976. **15**(4): p. 787-92.
64. Nowack, H., B.R. Olsen, and R. Timpl, *Characterization of amino-terminal segment in type-3 procollagen.* European Journal of Biochemistry, 1976. **70**(1): p. 205-216.
65. Brandt, A., et al., *Complete amino-acid-sequence of the N-terminal extension of calf skin type-III procollagen.* Biochemical Journal, 1984. **219**(2): p. 625-634.
66. Bruckner, P., et al., *3 conformationally distinct domains in amino-terminal segment of type-III procollagen and its rapid triple helix reversible coil transition.* European Journal of Biochemistry, 1978. **90**(3): p. 595-603.
67. Glanville, R.W. and P.P. Fietzek, *Amino acid sequence of the N-terminal non-triple helical cross link region of type III collagen.* FEBS Letters, 1976. **71**(1): p. 99-102.
68. Alakokko, L., et al., *Structure of cDNA clones coding for the entire prepro-alpha-1(III) chain of human type-III procollagen - differences in protein-structure from type-i procollagen and conservation of codon preferences.* Biochemical Journal, 1989. **260**(2): p. 509-516.
69. Wikipedia. *Edman degradation* [cited 2018 03/01/2018]; Available from: https://en.wikipedia.org/w/index.php?title=Edman_degradation&oldid=809376943.
70. Niemela, O., *Radioimmunoassays for type-III procollagen amino-terminal peptides in humans.* Clinical Chemistry, 1985. **31**(8): p. 1301-1304.
71. Rohde, H., P. Bruckner, and R. Timpl, *Immunochemical properties of the aminopropeptide of procollagen type-III.* European Journal of Biochemistry, 1983. **135**(2): p. 197-202.
72. Niemela, O., et al., *Purification and characterization of the N-terminal propeptide of human type-III procollagen.* Biochemical Journal, 1985. **232**(1): p. 145-150.
73. Niemela, O., et al., *Heterogeneity of the antigens related to the aminoterminal propeptide of type-III procollagen in human-serum.* Clinica Chimica Acta, 1982. **124**(1): p. 39-44.
74. Cantin, A.M., R. Boileau, and R. Bégin, *Increased Procollagen III Aminoterminal Peptide-related Antigens and Fibroblast Growth Signals in the Lungs of Patients with Idiopathic Pulmonary Fibrosis.* American Review of Respiratory Disease, 1988. **137**(3): p. 572-578.
75. Haukipuro, K., et al., *Aminoterminal propeptide of type III procollagen in healing wound in humans.* Ann Surg, 1987. **206**(6): p. 752-6.
76. Bowness, J.M., J.E. Folk, and R. Timpl, *Identification of a substrate site for liver transglutaminase on the aminopropeptide of type III collagen.* J Biol Chem, 1987. **262**(3): p. 1022-4.
77. Jensen, L.T., et al., *Fate of circulating amino-terminal propeptide of type-III procollagen in conscious pigs.* American Journal of Physiology, 1993. **265**(1): p. R139-R145.
78. Jeffers, L.J., et al., *Procollagen-III peptide and chronic viral C-hepatitis.* American Journal of Gastroenterology, 1995. **90**(9): p. 1437-1440.
79. Ramadori, G., et al., *Serum hyaluronate and type-III procollagen aminoterminal propeptide concentration in chronic liver-disease - relationship to cirrhosis and disease-activity.* European Journal of Clinical Investigation, 1991. **21**(3): p. 323-330.
80. Bentsen, K.D., et al., *Serum aminoterminal procollagen type-III peptide in acute viral-hepatitis - a long-term follow-up-study.* Liver, 1987. **7**(2): p. 96-105.
81. Lichtinghagen, R., et al., *The Enhanced Liver Fibrosis (ELF) score: Normal values, influence factors and proposed cut-off values.* Journal of Hepatology, 2013. **59**(2): p. 236-242.
82. Trivedi, P., et al., *Growth velocity, growth-hormone therapy, and serum concentrations of the amino-terminal propeptide of type-III procollagen.* Journal of Pediatrics, 1989. **114**(2): p. 225-230.

83. Rogol, A.D., *Growth hormone and the adolescent athlete: What are the data for its safety and efficacy as an ergogenic agent?* Growth Hormone & Igf Research, 2009. **19**(4): p. 294-299.
84. Hoofnagle, A.N. and M.H. Wener, *The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry.* Journal of Immunological Methods, 2009. **347**(1-2): p. 3-11.
85. Burchardt, E.R., et al., *Monoclonal antibody and assay for detecting PIIINP*. 2009, Google Patents.
86. Rohde, H., et al., *Radioimmunoassay for type-III procollagen peptide and its application to human-liver disease.* European Journal of Clinical Investigation, 1979. **9**(6): p. 451-459.
87. Unterweger, M.P., D.D. Hoppes, and F.J. Schima, *New and revised half-life measurements results.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1992. **312**(1): p. 349-352.
88. Haudenschild, D.R., et al., *High abundant protein removal from rodent blood for biomarker discovery.* Biochemical and Biophysical Research Communications, 2014. **455**(1-2): p. 84-89.
89. Echan, L.A., et al., *Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma.* Proteomics – Clinical Applications, 2005. **5**(13): p. 3292-3303.
90. Polson, C., et al., *Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography–tandem mass spectrometry.* Journal of Chromatography B, 2003. **785**(2): p. 263-275.
91. Greening, D.W. and R.J. Simpson, *A centrifugal ultrafiltration strategy for isolating the low-molecular weight (≤25K) component of human plasma proteome.* Journal of Proteomics, 2010. **73**(3): p. 637-648.
92. Kullolli, M., W.S. Hancock, and M. Hincapie, *Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins.* Journal of Separation Science, 2008. **31**(14): p. 2733-2739.
93. Winther, B., et al., *Immuno-capture as ultimate sample cleanup in LC-MS/MS determination of the early stage biomarker ProGRP.* Journal of Separation Science, 2009. **32**(17): p. 2937-2943.
94. Dittrich, J., et al., *Sample preparation strategies for targeted proteomics via proteotypic peptides in human blood using liquid chromatography tandem mass spectrometry.* Proteomics – Clinical applications, 2015. **9**(1-2): p. 5-16.
95. Ewles, M. and L. Goodwin, *Bioanalytical approaches to analyzing peptides and proteins by LC-MS/MS.* Bioanalysis, 2011. **3**(12): p. 1379-1397.
96. Van den Broek, I., et al., *Current trends in mass spectrometry of peptides and proteins: Application to veterinary and sports-doping control.* Mass Spectrometry Reviews, 2015. **34**(6): p. 571-594.
97. Van den Broek, I., et al., *Quantitative bioanalysis of peptides by liquid chromatography coupled to (tandem) mass spectrometry.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2008. **872**(1-2): p. 1-22.
98. Dole, M., et al., *Molecular Beams of Macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.
99. Iribarne, J.V. and B.A. Thomson, *On the evaporation of small ions from charged droplets.* The Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.
100. Kebarle, P., *A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry.* Journal of Mass Spectrometry, 2000. **35**(7): p. 804-817.
101. El-Aneed, A., A. Cohen, and J. Banoub, *Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers.* Applied Spectroscopy Reviews, 2009. **44**(3): p. 210-230.

102. Baldwin, M.A., *Mass Spectrometers for the Analysis of Biomolecules*, in *Methods in Enzymology*. 2005, Academic Press. p. 3-48.
103. Steen, H. and M. Mann, *The abc's (and xyz's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.
104. Hoffman, E., Strobant, V., *Mass Spectrometry: Principles and Applications, John Wiley & Sons (Third Edition), New York* 2002.
105. Watson, J.T., Sparkman, O. D., *Introduction to Mass Spectrometry.* John Wiley & Sons, Ltd., Southern Gate, Chichester (third edition), 2008.
106. *Quadrupole ion trap*. 03/01/2018]; Available from: http://nptel.ac.in/courses/102103044/module2/lec12/5.html.
107. Clauser, K.R., P. Baker, and A.L. Burlingame, *Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching.* Analytical Chemistry, 1999. **71**(14): p. 2871-2882.
108. Scigelova, M. and A. Makarov, *Orbitrap mass analyzer--overview and applications in proteomics.* Proteomics, 2006. **6 Suppl 2**: p. 16-21.
109. Hipple, J.A., H. Sommer, and H.A. Thomas, *A Precise Method of Determining the Faraday by Magnetic Resonance.* Physical Review, 1949. **76**(12): p. 1877-1878.
110. Paizs, B. and S. Suhai, *Fragmentation pathways of protonated peptides.* Mass Spectrometry Reviews, 2005. **24**(4): p. 508-548.
111. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass-spectra of peptides.* Biomedical Mass Spectrometry, 1984. **11**(11): p. 601-601.
112. Biemann, K., *Contributions of mass-spectrometry to peptide and protein-structure.* Biomedical and Environmental Mass Spectrometry, 1988. **16**(1-12): p. 99-111.
113. Dongre, A.R., et al., *Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model.* Journal of the American Chemical Society, 1996. **118**(35): p. 8365-8374.
114. Boyd, R.K., C. Basic, and R.A. Bethem, *Tools of the Trade V. Mass Analyzers for Quantitation: Separation of Ions by m/z Values*, in *Trace Quantitative Analysis by Mass Spectrometry*. 2008, John Wiley & Sons, Ltd. p. 245-343.
115. Liu, W., et al., *Recombinant human collagen for tissue engineered corneal substitutes.* Biomaterials, 2008. **29**(9): p. 1147-1158.
116. Yamada, K., et al., *Fabrication of cultured oral gingiva by tissue engineering techniques without materials of animal origin.* Journal of Periodontology, 2006. **77**(4): p. 672-677.
117. Liu, Y., et al., *Properties of porcine and recombinant human collagen matrices for optically clear tissue engineering applications.* Biomacromolecules, 2006. **7**(6): p. 1819-1828.
118. Baez, J., D. Olsen, and J.W. Polarek, *Recombinant microbial systems for the production of human collagen and gelatin.* Applied Microbiology and Biotechnology, 2005. **69**(3): p. 245-252.
119. Stawikowski, M. and G.B. Fields, *Introduction to Peptide Synthesis.* Current protocols in protein science / editorial board, John E. Coligan ... [et al.], 2002: p. Unit-18.1.
120. Robinson, A.B. and L.R. Robinson, *Distribution of glutamine and asparagine residues and their near neighbors in peptides and proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(20): p. 8880-8884.
121. Li, X., C. Lin, and P.B. O'Connor, *Glutamine Deamidation: Differentiation of Glutamic Acid and gamma-Glutamic Acid in Peptides by Electron Capture Dissociation.* Analytical Chemistry, 2010. **82**(9): p. 3606-3615.

122. Krokhin, O.V., et al., *Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis.* Analytical Chemistry, 2006. **78**(18): p. 6645-6650.
123. Rivers, J., et al., *Asparagine deamidation and the role of higher order protein structure.* Journal of Proteome Research, 2008. **7**(3): p. 921-927.
124. Liu, Y.D., et al., *N-terminal Glutamate to Pyroglutamate Conversion in Vivo for Human IgG2 Antibodies.* Journal of Biological Chemistry, 2011. **286**(13): p. 11211-11217.
125. Schilling, S., C. Wasternack, and H.U. Demuth, *Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution.* Biological Chemistry, 2008. **389**(8): p. 983-991.
126. Horlein, D., et al., *Amino-acid sequence of the aminoterminal segment of dermatosparactic calf-skin procollagen type-I.* European Journal of Biochemistry, 1979. **99**(1): p. 31-38.
127. Rohde, H., et al., *Amino-acid sequence of the N-terminal non-collagenous segment of dermatosparactic sheep procollagen type-I.* Biochemical Journal, 1979. **179**(3): p. 631-642.
128. Bella, J., B. Brodsky, and H.M. Berman, *Hydration structure of a collagen peptide.* Structure, 1995. **3**(9): p. 893-906.
129. Nelson, D.L. and M.M. Cox, *Amino Acids, Peptides and Proteins*, in *Lehninger Principles of Biochemistry*. 2013, Palgrave Macmillan. p. 127-133.
130. Ramachan, G., M. Bansal, and R. Bhatnaga, *Hypothesis on role of hydroxyproline in stabilizing collagen structure.* Biochimica Et Biophysica Acta, 1973. **322**(1): p. 166-171.
131. Varki, A. and N. Sharon, *Historical background and overview*, in *Essentials of Glycobiology*, C.R. Varki A, Esko JD, et al., editors., Editor. 2009, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
132. Wong, C.H., *Protein glycosylation: New challenges and opportunities.* Journal of Organic Chemistry, 2005. **70**(11): p. 4219-4225.
133. Zhong, X. and W. Somer, *Recent Advances in Glycosylation Modifications in the Context of Therapeutic Glycoproteins*, in *Integrative Proteomics*, D.H.-C. Leung, Editor. 2012.
134. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annual Review of Biochemistry, 2004. **73**: p. 1019-1049.
135. Schwarz, F. and M. Aebi, *Mechanisms and principles of N-linked protein glycosylation.* Current Opinion in Structural Biology, 2011. **21**(5): p. 576-582.
136. Van den Steen, P., et al., *Concepts and principles of O-linked glycosylation.* Critical Reviews in Biochemistry and Molecular Biology, 1998. **33**(3): p. 151-208.
137. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Molecular & Cellular Proteomics, 2004. **3**(6): p. 608-614.
138. Tabb, D.L., et al., *Influence of basic residue content on fragment ion peak intensities in low-energy - collision-induced dissociation spectra of peptides.* Analytical Chemistry, 2004. **76**(5): p. 1243-1248.
139. Walmsley, S.J., et al., *Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin As a Significant Source of Variability in Proteomics.* Journal of Proteome Research, 2013. **12**(12): p. 5666-5680.
140. Vandermarliere, E., M. Mueller, and L. Martens, *Getting intimate with trypsin, the leading protease in proteomics.* Mass Spectrometry Reviews, 2013. **32**(6): p. 453-465.
141. Keil, B., *Specificity of proteolysis*. 1992: Springer-Verlag Berlin Heidelberg.
142. Perona, J.J. and C.S. Craik, *Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold.* Journal of Biological Chemistry, 1997. **272**(48): p. 29987-29990.

143. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J. Walker, Editor. 2005, Humana Press. p. 571-607.
144. Mass Spectrometry Facility, U.S.C.F., *Protein Prospector*. 1996-2014, The Regents of the University of California: The University of California.
145. Hervey, W.J.I.V., M.B. Strader, and G.B. Hurst, *Comparison of digestion protocols for microgram quantities of enriched protein samples.* Journal of Proteome Research, 2007. **6**(8): p. 3054-3061.
146. Tarentino, A.L., et al., *Molecular-cloning and sequence-analysis of flavastacin - an o-glycosylated prokaryotic zinc metalloendopeptidase.* Archives of Biochemistry and Biophysics, 1995. **319**(1): p. 281-285.
147. Drapeau, G.R., J. Houmard, and Y. Boily, *Purification and properties of an extracellular protease of staphylococcus-aureus.* Journal of Biological Chemistry, 1972. **247**(20): p. 6720-6726.
148. Piszkiew.D, M. Landon, and E.L. Smith, *Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations.* Biochemical and Biophysical Research Communications, 1970. **40**(5): p. 1173-1178.
149. Tabb, D.L., et al., *Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides.* Anal Chem, 2003. **75**(5): p. 1155-1163.
150. Knudsen, C.S., L. Heickendorff, and E. Nexo, *Measurement of amino terminal propeptide of type III procollagen (PIIINP) employing the ADVIA Centaur platform. Validation, reference interval and comparison to UniQ RIA.* Clinical Chemistry and Laboratory Medicine, 2014. **52**(2): p. 237-241.
151. Watson, J.T. and O.D. Sparkman, *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*. 2008: Wiley.
152. Emmett, M.R. and R.M. Caprioli, *Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins.* Journal of the American Society for Mass Spectrometry, 1994. **5**(7): p. 605-613.
153. Meiring, H.D., et al., *Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis.* Journal of Separation Science, 2002. **25**(9): p. 557-568.
154. *Practical considerations in enhancing LC–MS sensitivity for therapeutic protein bioanalysis.* Bioanalysis, 2017. **9**(18): p. 1353-1356.
155. Waters. [cited 2017 September 18]; Available from: http://www.waters.com/waters/en_GB/Nano--and-Microflow-LC-MS-Columns/nav.htm?cid=134778930&locale=en_GB.
156. Zhang, M., et al., *A sensitive, high-throughput and robust Trapping Micro-LC-MS strategy for quantification of biomarkers and antibody biotherapeutics.* Analytical Chemistry, 2017.
157. Shen, Y., et al., *Ultrasensitive Proteomics Using High-Efficiency On-Line Micro-SPE-NanoLC-NanoESI MS and MS/MS.* Analytical Chemistry, 2004. **76**(1): p. 144-154.
158. Kucera, P., *Microcolumn high-performance liquid chromatography*. Vol. 28. 2000: Elsevier.
159. Anderson, N.L. and N.G. Anderson, *The Human Plasma Proteome: History, Character, and Diagnostic Prospects.* Molecular & Cellular Proteomics, 2002. **1**(11): p. 845-867.
160. Tirumalai, R.S., et al., *Characterization of the low molecular weight human serum proteome.* Mol Cell Proteomics, 2003. **2**(10): p. 1096-103.
161. Björhall, K., T. Miliotis, and P. Davidsson, *Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples.* Proteomics – Clinical Applications, 2005. **5**(1): p. 307-317.
162. Zougman, A., et al., *Integrated Analysis of the Cerebrospinal Fluid Peptidome and Proteome.* Journal of Proteome Research, 2008. **7**(1): p. 386-399.
163. Merrell, K., et al., *Analysis of Low-Abundance, Low-Molecular-Weight Serum Proteins Using Mass Spectrometry*. Vol. 15. 2005. 238-48.

164. Liu, G., et al., *A Novel and Cost Effective Method of Removing Excess Albumin from Plasma/Serum Samples and Its Impacts on LC-MS/MS Bioanalysis of Therapeutic Proteins.* Analytical Chemistry, 2014. **86**(16): p. 8336-8343.
165. Frand, A.R., J.W. Cuozzo, and C.A. Kaiser, *Pathways for protein disulphide bond formation.* Trends in Cell Biology, 2000. **10**(5): p. 203-210.
166. Glanville, R.W., H. Allmann, and P.P. Fietzek, *Cysteine residues mark N-terminal triple-helical-non-triple-helical junction in typeIII collagen.* Hoppe-Seylers Zeitschrift Fur Physiologische Chemie, 1976. **357**(11): p. 1663-1665.
167. Quinlan, G.J., G.S. Martin, and T.W. Evans, *Albumin: Biochemical properties and therapeutic potential.* Hepatology, 2005. **41**(6): p. 1211-1219.
168. Castellanos, M.M. and C.M. Colina, *Molecular Dynamics Simulations of Human Serum Albumin and Role of Disulfide Bonds.* The Journal of Physical Chemistry B, 2013. **117**(40): p. 11895-11905.
169. Kay, R., et al., *Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis.* Rapid Communications in Mass Spectrometry, 2008. **22**(20): p. 3255-3260.
170. Lollo, B.A., et al., *Improved two-dimensional gel electrophoresis representation of serum proteins by using ProtoClearTM.* Electrophoresis, 1999. **20**(4-5): p. 854-859.
171. Ahmed, N., et al., *An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum.* Proteomics – Clinical Applications, 2003. **3**(10): p. 1980-1987.
172. Ghosh, R. and Z.F. Cui, *Protein purification by ultrafiltration with pre-treated membrane.* Journal of Membrane Science, 2000. **167**(1): p. 47-53.
173. *Protein Purification and Handling*. Available from: https://laboratory.pall.com/content/dam/pall/laboratory/literature-library/non-gated/id-35497.pdf.
174. Harper, R.G., et al., *Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry.* Electrophoresis, 2004. **25**(9): p. 1299-1306.
175. Gianazza, E. and P. Arnaud, *A general method for fractionation of plasma proteins. Dye-ligand affinity chromatography on immobilized Cibacron blue F3-GA.* Biochemical Journal, 1982. **201**(1): p. 129-136.
176. Wang, Y.Y., P. Cheng, and D.W. Chan, *A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis.* Proteomics, 2003. **3**(3): p. 243-8.
177. Thompson, S.T., K.H. Cass, and E. Stellwagen, *Blue dextran-sepharose: an affinity column for the dinucleotide fold in proteins.* Proc Natl Acad Sci U S A, 1975. **72**(2): p. 669-72.
178. Hoofnagle, A.N., et al., *Quantification of Thyroglobulin, a Low-Abundance Serum Protein, by Immunoaffinity Peptide Enrichment and Tandem Mass Spectrometry.* Clinical Chemistry, 2008. **54**(11): p. 1796-1804.
179. Tucholska, M., et al., *Endogenous peptides from biophysical and biochemical fractionation of serum analyzed by matrix-assisted laser desorption/ionization and electrospray ionization hybrid quadrupole time-of-flight.* Analytical Biochemistry, 2007. **370**(2): p. 228-245.
180. Orvisky, E., et al., *Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma.* Proteomics – Clinical Applications, 2006. **6**(9): p. 2895-2902.
181. *Minimum criteria for chromatographic-mass spectrometric confirmation of the identity of analytes for doping control purposes.*, in *WADA Technical Document – TD2015IDCR*. 2015.
182. Hoofnagle, A.N., et al., *Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry–Based Assays.* Clinical Chemistry, 2016. **62**(1): p. 48-69.

183. Zhang, G., et al., *Protein Quantitation Using Mass Spectrometry*, in *Computational Biology*, D. Fenyö, Editor. 2010, Humana Press: Totowa, NJ. p. 211-222.
184. Nilsson, B.L., M.B. Soellner, and R.T. Raines, *Chemical Synthesis of Proteins.* Annual Review of Biophysics and Biomolecular Structure, 2005. **34**(1): p. 91-118.
185. Palomo, J.M., *Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides.* RSC Advances, 2014. **4**(62): p. 32658-32672.
186. Kaiser, E., et al., *Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides.* Analytical Biochemistry, 1970. **34**(2): p. 595-598.
187. Hancock, W.S. and J.E. Battersby, *A new micro-test for the detection of incomplete coupling reactions in solid-phase peptide synthesis using 2,4,6-trinitrobenzene-sulphonic acid.* Analytical Biochemistry, 1976. **71**(1): p. 260-264.
188. Liebler, D.C. and L.J. Zimmerman, *Targeted Quantitation of Proteins by Mass Spectrometry.* Biochemistry, 2013. **52**(22): p. 3797-3806.
189. Prinsen, H.C.M.T., et al., *Rapid quantification of underivatized amino acids in plasma by hydrophilic interaction liquid chromatography (HILIC) coupled with tandem mass-spectrometry.* Journal of Inherited Metabolic Disease, 2016. **39**(5): p. 651-660.

Appendix I

(Mascot reports for proteins in HSA depleted matrix)

Mascot Search Results

User : Discoverer_Anyone
Email :
Search title : 090516_DD_002_Node:2
MS data file : File Name: 090516_DD_002.raw
Database : uniprot_sprot sprot_130220 (539165 sequences; 191456931 residues)
Timestamp : 12 May 2016 at 13:27:30 GMT
Protein hits : P00738 HPT_HUMAN Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1
Q5NVH5 ALBU_PONAB Serum albumin OS=Pongo abelii GN=ALB PE=2 SV=2
P00734 THRB_HUMAN Prothrombin OS=Homo sapiens GN=F2 PE=1 SV=2
P0DJG0 APOA1_PANTR Apolipoprotein A-I OS=Pan troglodytes GN=APOA1 PE=1 SV=1
P02656 APOC3_HUMAN Apolipoprotein C-III OS=Homo sapiens GN=APOC3 PE=1 SV=1
P01834 IGKC_HUMAN Ig kappa chain C region OS=Homo sapiens GN=IGKC PE=1 SV=1
P0CG04 LAC1_HUMAN Ig lambda-1 chain C regions OS=Homo sapiens GN=IGLC1 PE=1 SV=1
P02671 FIBA_HUMAN Fibrinogen alpha chain OS=Homo sapiens GN=FGA PE=1 SV=2
P0DJG2 APOA2_GORGO Apolipoprotein A-II OS=Gorilla gorilla gorilla GN=APOA2 PE=1 SV=1
P0C0L4 CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=1
P0DKU9 APOE_COLGU Apolipoprotein E OS=Colobus guereza GN=APOE PE=3 SV=1
Q9GLM6 APOE_HYLLA Apolipoprotein E OS=Hylobates lar GN=APOE PE=3 SV=1
P00761 TRYP_PIG Trypsin OS=Sus scrofa PE=1 SV=1
Q9N2D0 FETUA_PANTR Alpha-2-HS-glycoprotein OS=Pan troglodytes GN=AHSG PE=3 SV=1
O67898 SYGB_AQUAE Glycine--tRNA ligase beta subunit OS=Aquifex aeolicus (strain VF5) GN=glyS PE=3 SV=1
Q8WZ42 TITIN_HUMAN Titin OS=Homo sapiens GN=TTN PE=1 SV=4
Q6CXW4 GCN5_KLULA Histone acetyltransferase GCN5 OS=Kluyveromyces lactis (strain ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37) GN=GCN5 PE=3 SV=1
P0DKU5 APOC2_COLGU Apolipoprotein C-II OS=Colobus guereza GN=APOC2 PE=3 SV=1
P19034 APOC2_BOVIN Apolipoprotein C-II OS=Bos taurus GN=APOC2 PE=1 SV=2
A6NMZ7 CO6A6_HUMAN Collagen alpha-6(VI) chain OS=Homo sapiens GN=COL6A6 PE=1 SV=2
Q73K81 SYL_TREDE Leucine--tRNA ligase OS=Treponema denticola (strain ATCC 35405 / CIP 103919 / DSM 14222) GN=leuS PE=3 SV=1
Q72KH8 END4_THET2 Probable endonuclease 4 OS=Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039) GN=nfo PE=3 SV=1
Q99NB9 SF3B1_MOUSE Splicing factor 3B subunit 1 OS=Mus musculus GN=Sf3b1 PE=1 SV=1
B8CME8 SYT_SHEPW Threonine--tRNA ligase OS=Shewanella piezotolerans (strain WP3 / JCM 13877) GN=thrS PE=3 SV=1
Q23551 UNC22_CAEEL Twitchin OS=Caenorhabditis elegans GN=unc-22 PE=1 SV=3

	uniprot_sprot	Decoy	False discovery rate
Peptide matches above identity threshold	22	6	27.27 %
Peptide matches above homology or identity threshold	35	8	22.86 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 34 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

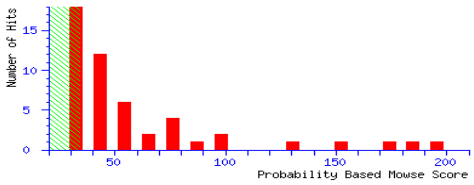


Figure 1: Proteins identified in double MWCO (30 kDa) depleted serum by Mascot

Mascot Search Results

User : Discoverer_Anyone
Email :
Search title : 090516_DD_004_Node:2
MS data file : File Name: 090516_DD_004.raw
Database : uniprot_sprot sprot_130220 (539165 sequences; 191456931 residues)
Timestamp : 12 May 2016 at 13:28:37 GMT
Protein hits : P0DJG0 APOA1_PANTR Apolipoprotein A-I OS=Pan troglodytes GN=APOA1 PE=1 SV=1
P02775 CXCL7_HUMAN Platelet basic protein OS=Homo sapiens GN=PPBP PE=1 SV=3
Q9GLM6 APOE_HYLLA Apolipoprotein E OS=Hylobates lar GN=APOE PE=3 SV=1
P0DKU9 APOE_COLGU Apolipoprotein E OS=Colobus guereza GN=APOE PE=3 SV=1
P02648 APOA1_CANFA Apolipoprotein A-I OS=Canis familiaris GN=APOA1 PE=1 SV=2
P0DKW8 APOE_SAIBB Apolipoprotein E (Fragment) OS=Saimiri boliviensis boliviensis GN=APOE PE=3 SV=1
P0C0L4 CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=1
P0DJG2 APOA2_GORGO Apolipoprotein A-II OS=Gorilla gorilla gorilla GN=APOA2 PE=1 SV=1
P02656 APOC3_HUMAN Apolipoprotein C-III OS=Homo sapiens GN=APOC3 PE=1 SV=1
P01024 CO3_HUMAN Complement C3 OS=Homo sapiens GN=C3 PE=1 SV=2
P06727 APOA4_HUMAN Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3
A5A6I6 TRFE_PANTR Serotransferrin OS=Pan troglodytes GN=TF PE=2 SV=1
Q28758 APOA4_PAPAN Apolipoprotein A-IV (Fragment) OS=Papio anubis GN=APOA4 PE=2 SV=1
B2KHZ4 HBB_RHIFE Hemoglobin subunit beta OS=Rhinolophus ferrumequinum GN=HBB PE=2 SV=1
P30002 TEGU_HHV6G Large tegument protein OS=Human herpesvirus 6A (strain GS) GN=U31 PE=3 SV=1
Q8RX22 MIP1_ARATH MND1-interacting protein 1 OS=Arabidopsis thaliana GN=MIP1 PE=1 SV=1
P01030 CO4_BOVIN Complement C4 (Fragments) OS=Bos taurus GN=C4 PE=1 SV=2
Q8C6K9 CO6A6_MOUSE Collagen alpha-6(VI) chain OS=Mus musculus GN=Col6a6 PE=1 SV=2
P08250 APOA1_CHICK Apolipoprotein A-I OS=Gallus gallus GN=APOA1 PE=1 SV=2
P00761 TRYP_PIG Trypsin OS=Sus scrofa PE=1 SV=1
A0K3L7 RPOB_BURCH DNA-directed RNA polymerase subunit beta OS=Burkholderia cenocepacia (strain HI2424) GN=rpoB PE=3 SV=1
P02654 APOC1_HUMAN Apolipoprotein C-I OS=Homo sapiens GN=APOC1 PE=1 SV=1
Q9I7U4 TITIN_DROME Titin OS=Drosophila melanogaster GN=sls PE=1 SV=3
P02036 HBB_SAISC Hemoglobin subunit beta OS=Saimiri sciureus GN=HBB PE=1 SV=2
O28552 DP2L_ARCFU DNA polymerase II large subunit OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=polC PE=3 SV=1

	uniprot_sprot	Decoy	False discovery rate
Peptide matches above identity threshold	32	10	31.25 %
Peptide matches above homology or identity threshold	45	13	28.89 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 33 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

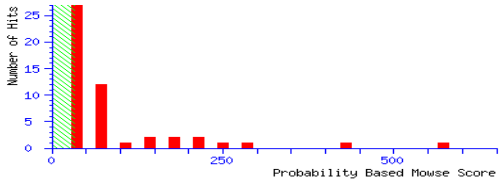


Figure 2: Proteins identified in protein precipitation and MWCO (30 kDa) depleted serum by Mascot

Mascot Search Results

```
User                : Discoverer_Anyone
Email               :
Search title        : 090516_DD_006_Node:2
MS data file        : File Name: 090516_DD_006.raw
Database            : uniprot_sprot sprot_130220 (539165 sequences; 191456931 residues)
Timestamp           : 12 May 2016 at 13:45:27 GMT
Protein hits        : P02768 ALBU_HUMAN Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2
                      P0DJG0 APOA1_PANTR Apolipoprotein A-I OS=Pan troglodytes GN=APOA1 PE=1 SV=1
                      Q5NVH5 ALBU_PONAB Serum albumin OS=Pongo abelii GN=ALB PE=2 SV=2
                      P01024 CO3_HUMAN Complement C3 OS=Homo sapiens GN=C3 PE=1 SV=2
                      P00738 HPT_HUMAN Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1
                      P0DJG2 APOA2_GORGO Apolipoprotein A-II OS=Gorilla gorilla gorilla GN=APOA2 PE=1 SV=1
                      P0CG05 LAC2_HUMAN Ig lambda-2 chain C regions OS=Homo sapiens GN=IGLC2 PE=1 SV=1
                      P02648 APOA1_CANFA Apolipoprotein A-I OS=Canis familiaris GN=APOA1 PE=1 SV=2
                      P0CG04 LAC1_HUMAN Ig lambda-1 chain C regions OS=Homo sapiens GN=IGLC1 PE=1 SV=1
                      P01834 IGKC_HUMAN Ig kappa chain C region OS=Homo sapiens GN=IGKC PE=1 SV=1
                      P00734 THRB_HUMAN Prothrombin OS=Homo sapiens GN=F2 PE=1 SV=2
                      P06727 APOA4_HUMAN Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3
                      P0C0L4 CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=1
                      P02656 APOC3_HUMAN Apolipoprotein C-III OS=Homo sapiens GN=APOC3 PE=1 SV=1
                      P02765 FETUA_HUMAN Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1
                      P08835 ALBU_PIG Serum albumin OS=Sus scrofa GN=ALB PE=1 SV=2
                      Q9GLM6 APOE_HYLLA Apolipoprotein E OS=Hylobates lar GN=APOE PE=3 SV=1
                      Q2UVX4 CO3_BOVIN Complement C3 OS=Bos taurus GN=C3 PE=1 SV=2
                      P01857 IGHG1_HUMAN Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1
                      P12387 CO3_CAVPO Complement C3 OS=Cavia porcellus GN=C3 PE=1 SV=2
                      P02671 FIBA_HUMAN Fibrinogen alpha chain OS=Homo sapiens GN=FGA PE=1 SV=2
                      P01009 A1AT_HUMAN Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3
                      P35527 K1C9_HUMAN Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3
                      P01591 IGJ_HUMAN Immunoglobulin J chain OS=Homo sapiens GN=IGJ PE=1 SV=4
                      Q5XLE4 ALBU_EQUAS Serum albumin OS=Equus asinus GN=ALB PE=1 SV=1
```

	uniprot_sprot	Decoy	False discovery rate
Peptide matches above identity threshold	72	13	18.06 %
Peptide matches above homology or identity threshold	99	22	22.22 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores ≥ 35 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

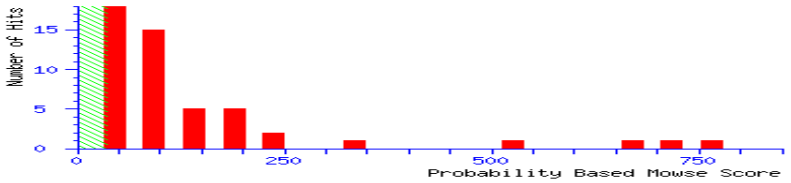


Figure 3: Proteins identified in MWCO (30 kDa) depleted serum by Mascot

Mascot Search Results

```
User                : Discoverer_Anyone
Email               :
Search title        : 090516_DD_008_Node:2
MS data file        : File Name: 090516_DD_008.raw
Database            : uniprot_sprot sprot_130220 (539165 sequences; 191456931 residues)
Timestamp           : 12 May 2016 at 15:04:16 GMT
Protein hits        : P02768 ALBU_HUMAN Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2
                      Q5NVH5 ALBU_PONAB Serum albumin OS=Pongo abelii GN=ALB PE=2 SV=2
                      P0DJG0 APOA1_PANTR Apolipoprotein A-I OS=Pan troglodytes GN=APOA1 PE=1 SV=1
                      P02787 TRFE_HUMAN Serotransferrin OS=Homo sapiens GN=TF PE=1 SV=3
                      A2V9Z4 ALBU_MACFA Serum albumin OS=Macaca fascicularis GN=ALB PE=2 SV=1
                      P68871 HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
                      P06727 APOA4_HUMAN Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3
                      P69905 HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
                      P0DJG2 APOA2_GORGO Apolipoprotein A-II OS=Gorilla gorilla gorilla GN=APOA2 PE=1 SV=1
                      Q9GLM6 APOE_HYLLA Apolipoprotein E OS=Hylobates lar GN=APOE PE=3 SV=1
                      Q6WN28 HBB_CALJA Hemoglobin subunit beta OS=Callithrix jacchus GN=HBB PE=2 SV=3
                      P02042 HBD_HUMAN Hemoglobin subunit delta OS=Homo sapiens GN=HBD PE=1 SV=2
                      P08835 ALBU_PIG Serum albumin OS=Sus scrofa GN=ALB PE=1 SV=2
                      Q6WN20 HBB_CALTO Hemoglobin subunit beta OS=Callicebus torquatus GN=HBB PE=2 SV=3
                      P0DKU9 APOE_COLGU Apolipoprotein E OS=Colobus guereza GN=APOE PE=3 SV=1
                      P02044 HBD_ATEGE Hemoglobin subunit delta OS=Ateles geoffroyi GN=HBD PE=1 SV=3
                      P26916 HBB_NASNA Hemoglobin subunit beta OS=Nasua nasua GN=HBB PE=1 SV=1
                      P02775 CXCL7_HUMAN Platelet basic protein OS=Homo sapiens GN=PPBP PE=1 SV=3
                      P02648 APOA1_CANFA Apolipoprotein A-I OS=Canis familiaris GN=APOA1 PE=1 SV=2
                      A6YF56 ALBU_MESAU Serum albumin OS=Mesocricetus auratus GN=ALB PE=1 SV=1
                      P02036 HBB_SAISC Hemoglobin subunit beta OS=Saimiri sciureus GN=HBB PE=1 SV=2
                      Q28758 APOA4_PAPAN Apolipoprotein A-IV (Fragment) OS=Papio anubis GN=APOA4 PE=2 SV=1
                      P01932 HBA_THEGE Hemoglobin subunit alpha OS=Theropithecus gelada GN=HBA PE=1 SV=1
                      P09420 HBA_SPECI Hemoglobin subunit alpha OS=Spermophilus citellus GN=HBA PE=1 SV=1
                      P10893 HBB_LUTLU Hemoglobin subunit beta OS=Lutra lutra GN=HBB PE=1 SV=1
```

	uniprot_sprot	Decoy	False discovery rate
Peptide matches above identity threshold	132	26	19.70 %
Peptide matches above homology or identity threshold	177	39	22.03 %

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 35 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

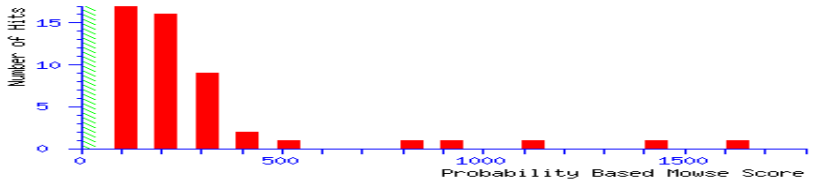


Figure 4: Proteins identified in protein precipitation depleted serum by Mascot